STRUCTURAL BIOINFORMATICS AND BIOPHYSICAL APPROACHES FOR UNDERSTANDING THE PLANT RESPONSES TO BIOTIC AND ABIOTIC STRESS

EDITED BY: Raul Antonio Sperotto, Maria Hrmova, Steffen P. Graether and Luis Fernando Saraiva Macedo Timmers
PUBLISHED IN: Frontiers in Plant Science

frontiers Research Topics

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

STRUCTURAL BIOINFORMATICS AND BIOPHYSICAL APPROACHES FOR UNDERSTANDING THE PLANT RESPONSES TO BIOTIC AND ABIOTIC STRESS

Topic Editors:
Raul Antonio Sperotto, Universidade do Vale do Taquari - Univates, Brazil
Maria Hrmova, University of Adelaide, Australia
Steffen P. Graether, University of Guelph, Canada
Luis Fernando Saraiva Macedo Timmers, Universidade do Vale do
Taquari - Univates, Brazil

Table of Contents

![frontiers | Frontiers in Plant Science]

Editorial: Structural bioinformatics and biophysical approaches for understanding the plant responses to biotic and abiotic stresses

Raul A. Sperotto[1,2]*, Maria Hrmova[3,4]*, Steffen P. Graether[5]* and Luis Fernando S. M. Timmers[1,6]*

[1]Graduate Program in Biotechnology, University of Taquari Valley – Univates, Lajeado, Brazil, [2]Graduate Program in Plant Physiology, Federal University of Pelotas, Pelotas, Brazil, [3]Faculty of Sciences, Engineering and Technology, University of Adelaide, Adelaide, SA, Australia, [4]School of Life Science, Huaiyin Normal University, Huaian, China, [5]Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada, [6]Graduate Program in Medical Sciences, University of Taquari Valley – Univates, Lajeado, Brazil

Editorial on the Research Topic
Structural bioinformatics and biophysical approaches for understanding the plant responses to biotic and abiotic stress

Plants are exposed to a variety of environmental conditions that negatively impact their physiology and yields (Rivero et al., 2022; Sharma et al., 2022; Zandalinas and Mittler, 2022). Several studies identified mechanisms involving genes, proteins, and metabolites that underlie plant responses to stress conditions (Chen et al., 2022; Hassan et al., 2022; Huang et al., 2022; Mittler et al., 2022; Yang et al., 2022; Zhan et al., 2022). Some of these molecules were used to improve plant responses to abiotic and biotic stresses (Li et al., 2022; Mahto et al., 2022; Wang and Komatsu, 2022; Zhao et al., 2022). However, the underlying structural and functional relationships of these molecular mechanisms require more research. Computational and biophysical approaches are viable options for analyses of target molecules to understand their interactions and dynamics that initiate biochemical and physiological responses of plants (Wan et al., 2015; Konda et al., 2018; Moffett and Shukla, 2018; Rayevsky et al., 2019; Jha et al., 2022). These responses in turn control plant tolerance and resistance to sub-optimal environmental conditions. Therefore, this Research Topic is aligned with the current research trends and provides an update on the advances in structural bioinformatics and biophysical approaches to understanding plant responses to biotic and abiotic stresses at the molecular level. Here, we highlight some of the topics from the following contributions.

Biotic stress in plants involves the complex regulatory networks of various sensory/signaling molecules (Tiwari et al., 2022), including proteins that participate in plant defense strategies. Three articles focused on the genetic characterization of proteins involved in biotic stress resistance. First, Zheng et al. identified protein-protein interactions (PPIs) between rice and the fungus *Magnaporthe oryzae*, a causative agent of rice blast, which is the most devastating disease affecting rice production. A global PPI network consisted of 2,018 interacting protein pairs involving 1,344 rice proteins, where 418 blast fungus proteins showed enrichment for blast resistance genes—these network-based predictions would now allow discoveries of blast resistance genes in rice. Next, Sati et al. investigated computationally the interaction of the PsoR protein of plant-beneficial *Pseudomonas* spp. with various root exudates to better understand inter-kingdom signaling between plants and plant growth-promoting rhizobacteria (PGPR). The PsoR protein produces a range of antifungal and insecticidal secondary metabolites, making them useful biocontrol agents and thus helping during plant growth. A total of 59 different low molecular mass phytochemicals were virtually screened with the PsoR protein by molecular docking. Two root exudates, saponarin and 2-benzoxazolinone (BOA), present in the root exudates of barley and wheat, respectively, showed suitable binding with PsoR, likely showing cross-kingdom interactions. Lastly, Yadav et al. performed a wide-range analysis of Ca^{2+}-sensing plant-specific calmodulin-like proteins (CMLs) in soybean to identify 41 true CMLs. Gene structural analysis and identifying conserved motifs and *cis*-acting elements in these targets strongly support their identity as the members of this family and involvement in stress responses. Further insights into the differential expression patterns of *GmCMLs* during *Spodoptera litura*-feeding, wounding, and signaling together with 3D structure prediction, identification of interacting domains, and docking of Ca^{2+} ions in *S. litura*-inducible GmCMLs provided evidence into their roles in Ca^{2+} signaling and plant defense during herbivory.

Two articles focused on the genetic characterization of proteins involved in abiotic stress tolerance and climate change and their impacts, suggesting potential candidates that could be targeted for plant breeding and genetic engineering. This is aimed at developing food crops that could thrive in deteriorating environmental conditions, and maintain or increase the crop yields (Ku et al., 2018). First, Arabia et al. outlined the role of the *Universal Stress Protein* (*USP*) gene family in rice (*Oryza sativa* L. ssp. *japonica*) to uncover 44 genes and their domain architecture that was key to the functional diversification under multi-stress environmental challenges. Next, Chen et al. described 33 *Brassica napus* L. *CONSTANS-LIKE* (*COL*) genes; these clustered into three subfamilies and exhibited conserved gene and protein structures, promoter motifs, and tissue-specific expression. The latter study also clarified the role of each

gene subfamily during growth and development, flowering, and circadian rhythms.

Understanding of how bioactive molecules interact with their targets can help to explain their structure-activity relationships. In particular, the computational analyses of protein-ligand complexes can help to unravel these key interactions (Anighoro, 2020), providing valuable insights into the plant response to stress conditions. Two articles applied X-ray crystallography, comparative modeling, and molecular dynamics simulations to evaluate protein activity and protein-ligand interactions. In the first study, Infantes et al. outlined the importance of a single amino acid residue variation (PYR/PYL/RCAR) in ABA receptors, which promotes the transition of the latch and gate loops to active conformation in ABA-dependent and independent modes. Furthermore, the authors identified that the niacin molecule can act as an *in vivo* antagonist of ABA. In the second work, Maia et al. investigated the enzyme specificities of PR-4 SUGARWINs in sugarcane defense against phytopathogens. Based on experimental and computational methods, the authors showed the crystal structure of SUGARWIN2, the first PR-4 protein lacking an RNase activity. In addition, the authors suggested that SUGARWIN2 is more relevant to sugarcane defense against pathogenic fungi, as SUGARWIN2 showed higher expression levels after the *Colletothricum falcatum* (Went) treatment compared to the expression of SUGARWIN1.

Two additional articles focused on lipid barriers, which play critical roles in plant biology by separating spaces and generating compartments, and as biological barriers protect cells and organelles from stress conditions (Guo et al., 2019). In the review by Murray and Graether, these authors looked at the biophysical mechanisms by which dehydrins, a group of plant abiotic stress proteins, protect plants from cold damage. Despite dehydrin's lack of structure and low sequence conservation, the lysine-rich K-segments stabilized model membranes by maintaining fluidity and preventing membrane fusion. The paper by Ingram et al. examined a very different kind of lipid barrier—nanobubbles. The authors characterized the recently discovered 10–1,000 nm bubbles surrounded by a lipid monolayer, using molecular dynamics simulations. They show that they were stable and avoided embolisms at radii of 35 nm and under pressures of 0 to −1.5 MPa. To reconcile the presence of nanobubbles with the water transpiration model, the authors proposed that nanobubbles continuously expand and collapse into smaller bubbles, and reuse the lipids.

Final comment

In summary, the work presented in this Research Topic documents how structural bioinformatics and computational

biophysical approaches could be effectively applied to study the structure-function relationships of plant proteins and the function of small molecules involved in biotic and abiotic stresses. These molecules are at the core of the plant defense mechanisms that allow plants to mitigate stresses and enhance plant abilities to respond to unfavorable environmental conditions.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Anighoro, A. (2020). Underappreciated chemical interactions in protein-ligand complexes. *Methods Mol. Biol.* 2114, 75–86. doi: 10.1007/978-1-0716-0282-9_5

Chen, J., Yang, S., Fan, B., Zhu, C., and Chen, Z. (2022). The mediator complex: a central coordinator of plant adaptive responses to environmental stresses. *Int. J. Mol. Sci.* 23, 6170. doi: 10.3390/ijms23116170

Guo, Q., Liu, L., and Barkla, B. J. (2019). Membrane lipid remodeling in response to salinity. *Int. J. Mol. Sci.* 20, 4264. doi: 10.3390/ijms20174264

Hassan, M. U., Mahmood, A., Awan, M. I., Maqbool, R., Aamer, M., Alhaithloul, H. A. S., et al. (2022). Melatonin-induced protection against plant abiotic stress: mechanisms and prospects. *Front Plant Sci.* 13, 902694. doi: 10.3389/fpls.2022.902694

Huang, Y., Wang, W., Yu, H., Peng, J., Hu, Z., and Chen, L. (2022). The role of 14-3-3 proteins in plant growth and response to abiotic stress. *Plant Cell Rep.* 41, 833–852. doi: 10.1007/s00299-021-02803-4

Jha, Y., Dehury, B., Kumar, S. P. J., Chaurasia, A., Singh, U. B., Yadav, M. K., et al. (2022). Delineation of molecular interactions of plant growth promoting bacteria induced β-1,3-glucanases and guanosine triphosphate ligand for antifungal response in rice: a molecular dynamics approach. *Mol. Biol. Rep.* 49, 2579–2589. doi: 10.1007/s11033-021-07059-5

Konda, A. K., Farmer, R., Soren, K. R., P. S. S., and Setti, A. (2018). Structural modelling and molecular dynamics of a multi-stress responsive WRKY TF-DNA complex towards elucidating its role in stress signalling mechanisms in chickpea. *J. Biomol. Struct. Dyn.* 36, 2279–2291. doi: 10.1080/07391102.2017.1349690

Ku, Y. S., Sintaha, M., Cheung, M. Y., and Lam, H. M. (2018). Plant hormone signaling crosstalks between biotic and abiotic stress responses. *Int. J. Mol. Sci.* 19, 3206. doi: 10.3390/ijms19103206

Li, H., Chen, H., Chen, L., and Wang, C. (2022). The role of hydrogen sulfide in plant roots during development and in response to abiotic stress. *Int. J. Mol. Sci.* 23, 1024. doi: 10.3390/ijms23031024

Mahto, R. K., Ambika, S. C., Chandana, B. S., Singh, R. K., Verma, S., Gahlaut, V., et al. (2022). Chickpea biofortification for cytokinin dehydrogenase via genome editing to enhance abiotic-biotic stress tolerance and food security. *Front. Genet.* 13, 900324. doi: 10.3389/fgene.2022.900324

Mittler, R., Zandalinas, S. I., Fichman, Y., and Van Breusegem, F. (2022). Reactive oxygen species signalling in plant stress responses. *Nat. Rev. Mol. Cell Biol.* doi: 10.1038/s41580-022-00499-2. [Epub ahead of print].

Moffett, A. S., and Shukla, D. (2018). Using molecular simulation to explore the nanoscale dynamics of the plant kinome. *Biochem. J.* 475, 905–921. doi: 10.1042/BCJ20170299

Rayevsky, A. V., Sharifi, M., Samofalova, D. A., Karpov, P. A., and Blume, Y. B. (2019). Structural and functional features of lysine acetylation of plant and animal tubulins. *Cell Biol. Int.* 43, 1040–1048. doi: 10.1002/cbin.10887

Rivero, R. M., Mittler, R., Blumwald, E., and Zandalinas, S. I. (2022). Developing climate-resilient crops: improving plant tolerance to stress combination. *Plant J.* 109, 373–389. doi: 10.1111/tpj.15483

Sharma, M., Kumar, P., Verma, V., Sharma, R., Bhargava, B., and Irfan, M. (2022). Understanding plant stress memory response for abiotic stress resilience: molecular insights and prospects. *Plant Physiol. Biochem.* 179, 10–24. doi: 10.1016/j.plaphy.2022.03.004

Tiwari, R. K., Lal, M. K., Kumar, R., Mangal, V., Altaf, M. A., Sharma, S., et al. (2022). Insight into melatonin-mediated response and signaling in the regulation of plant defense under biotic stress. *Plant Mol. Biol.* 109, 385–399. doi: 10.1007/s11103-021-01 202-3

Wan, H., Chang, S., Hu, J. P., Tian, Y. X., and Tian, X. H. (2015). Molecular dynamics simulations of ternary complexes: comparisons of LEAFY protein binding to different DNA Motifs. *J. Chem. Inf. Model.* 55, 784–794. doi: 10.1021/ci500705j

Wang, X., and Komatsu, S. (2022). The role of phytohormones in plant response to flooding. *Int. J. Mol. Sci.* 23, 6383. doi: 10.3390/ijms23126383

Yang, X., Jia, Z., Pu, Q., Tian, Y., Zhu, F., and Liu, Y. (2022). ABA mediates plant development and abiotic stress via alternative splicing. *Int. J. Mol. Sci.* 23, 3796. doi: 10.3390/ijms23073796

Zandalinas, S. I., and Mittler, R. (2022). Plant responses to multifactorial stress combination. *New Phytol.* 234, 1161–1167. doi: 10.1111/nph.18087

Zhan, X., Chen, Z., Chen, R., and Shen, C. (2022). Environmental and genetic factors involved in plant protection-associated secondary metabolite biosynthesis pathways. *Front. Plant Sci.* 13, 877304. doi: 10.3389/fpls.2022.877304

Zhao, H., Li, Z., Wang, Y., Wang, J., Xiao, M., Liu, H., et al. (2022). Cellulose synthase-like protein OsCSLD4 plays an important role in the response of rice to salt stress by mediating abscisic acid biosynthesis to regulate osmotic stress tolerance. *Plant Biotechnol. J.* 20, 468–484. doi: 10.1111/pbi.13729

Check for
updates

Predicting Protein–Protein Interactions Between Rice and Blast Fungus Using Structure-Based Approaches

Cunjian Zheng, Yuan Liu, Fangnan Sun, Lingxia Zhao* and Lida Zhang*

Department of Plant Science, School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai, China

Rice blast, caused by the fungus *Magnaporthe oryzae*, is the most devastating disease affecting rice production. Identification of protein–protein interactions (PPIs) is a critical step toward understanding the molecular mechanisms underlying resistance to blast fungus in rice. In this study, we presented a computational framework for predicting plant–pathogen PPIs based on structural information. Compared with the sequence-based methods, the structure-based approach showed to be more powerful in discovering new PPIs between plants and pathogens. Using the structure-based method, we generated a global PPI network consisted of 2,018 interacting protein pairs involving 1,344 rice proteins and 418 blast fungus proteins. The network analysis showed that blast resistance genes were enriched in the PPI network. The network-based prediction enabled systematic discovery of new blast resistance genes in rice. The network provided a global map to help accelerate the identification of blast resistance genes and advance our understanding of plant–pathogen interactions.

Keywords: protein-protein interactions, rice, blast fungus (*Magnaporthe oryzae*), protein structure, machine learning

INTRODUCTION

Rice blast, caused by the fungus *Magnaporthe oryzae*, is the most devastating disease affecting rice production. Due to the availability of both genome sequences and the accessibility of molecular genetic tools, the pathosystem between rice and blast fungus has become a model system for studying plant–pathogen interactions (Dean et al., 2005; International Rice Genome Sequencing Project., 2005). Although the molecular mechanisms of the plant immune system have been extensively investigated over the past decade, many aspects of the overall resistance picture remain poorly understood (Meng et al., 2019).

Protein–protein interactions (PPIs) play a critical role in molecular recognition between plants and pathogens. Identification of these PPIs is important for understanding the underlying molecular mechanisms against pathogen infection in plants. Experimental methods have been used to identify plant–pathogen PPIs (Mukhtar et al., 2011; Weßling et al., 2014; Cao et al., 2019), but the available interaction data are still far from depicting global maps of plant–pathogen interactions (Ammari et al., 2016). Only a few experimentally verified PPIs between rice and blast fungus have been reported in the individual studies, which is insufficient to elucidate the molecular mechanisms leading to disease resistance in rice (Kanzaki et al., 2012; Cesari et al., 2013).

To complement experimental methods for identifying PPIs, many computational methods have been developed to accelerate the discovery of PPIs (Tanwar and George Priya Doss, 2018). Most available computational methods such as interolog mapping (Matthews et al., 2001), domain-based inference (Deng et al., 2002), gene fusion (Marcotte et al., 1999), phylogenetic similarity (Pellegrini et al., 1999), and gene co-expression [28] are primarily focused on determining PPIs within a single organism. Some of these methods, such as interolog and domain-based inference, have also been applied to the interspecies PPI field and lead to the discovery of important biological insights in plant–pathogen interactions (Li et al., 2012; Sahu et al., 2014; Yang et al., 2019).

Recently, computational methods using structural information to predict PPIs have gained much attention due to the rapid growth of protein three-dimensional (3D) structures (Zhang et al., 2012; Burley et al., 2017). Rather than sequence-based methods, structure-based approaches could reveal the structural details of protein interactions (Mariano and Wuchty, 2017). However, the structure-based methods require the 3D characterization of each protein and therefore suffer from low coverage of the proteome. Predicting protein interactions based on homology-modeled structures might be a solution to this problem, which enabled the use of protein structural information on a genome-wide scale (Zhang et al., 2016; Liu et al., 2017; Zhao et al., 2019).

In this study, we presented a computational framework for predicting plant–pathogen PPIs based on structural information. Performance assessment showed that the structure-based method was powerful in discovering PPIs between plants and pathogens. Furthermore, we used the structure-based method to generate a global PPI network between rice and blast fungus proteins, which provided a valuable reference for systems understanding of plant and pathogen interactions.

MATERIALS AND METHODS

Data Sources

All experimentally verified host–pathogen interactions were collected from the Host–Pathogen Interaction Database (https://hpidb.igbb.msstate.edu/index.html) (Ammari et al., 2016). A total of 10,148 host–pathogen PPIs with homology models and experimental structures were used as the positive reference dataset, while the interacting protein pairs between hosts and pathogens were randomly shuffled to form the negative reference dataset.

The genome of *M. oryzae* strain 70–15 was downloaded from Ensembl Genomes (http://fungi.ensembl.org/Magnaporthe_oryzae/Info/Index), and the genome of *Oryza sativa* ssp. *japonica* cv. Nipponbare was downloaded from the MSU Rice Genome Annotation Project Database (http://rice.plantbiology.msu.edu/). A total of 38,864 non-transposable element protein sequences were identified in the rice genome.

Identification of Membrane and Secreted Proteins in Blast Fungus

The membrane proteins containing one or more transmembrane helices were predicted by using TMHMM (Möller et al., 2001). The putatively secreted proteins were identified when the protein containing a signal peptide was predicted by using SignalP-5.0 (Almagro Armenteros et al., 2019) and the extracellular localization of the protein was predicted by using WoLF PSORT (Horton et al., 2007).

Homology Modeling of Protein Structures

Homology models of proteins were built by using ModPipe (Pieper et al., 2014). The homology structure with the highest ModPipe quality score was selected for each protein according to the previously described criteria (Zhang et al., 2016; Liu et al., 2017).

Structural Template for the Interaction Model

A total of 157,771 protein complexes involving 328,671 chains were collected from the Protein Data Bank (PDB) (Burley et al., 2017). The chain–chain binary interfaces of protein complexes were generated by PIBASE with an interatomic distance cutoff of 6.05 Å (Davis and Sali, 2005).

Structure-Based Features

Structural alignment was used to find the closest PDB chains of homology models using TM-align with the cutoff score of 0.4 (Zhang and Skolnick, 2005). The interaction model of protein pair was created by superimposing the homology structures on their corresponding chains in the closest PDB template complex. Four structural features, including structural similarity (i.e., TM-score) and structural distance (i.e., root mean square deviation, RMSD) between protein homology models and their corresponding chains in the template, as well as the number and fraction of interacting residue pairs in the template that were preserved in the interaction model, were calculated for the prediction of PPIs. The detailed method for structural features refers to the previous study (Zhang et al., 2016).

Prediction of Interolog and Domain-Based PPIs

The potential PPIs were predicted using interolog mapping. Each protein was blasted against the experimentally determined PPI datasets to identify homologs with an *E*-value of $<10^{-5}$, a sequence identity of $>45\%$, and an aligned sequence coverage of $>50\%$. The experimentally determined PPIs were derived from the BioGRID (Oughtred et al., 2021), IntAct (Orchard et al., 2014), MINT (Calderone et al., 2020), DIP (Salwinski et al., 2004), and BIND (Alfarano et al., 2005) databases (**Supplementary Table 1**).

The domains of each protein were identified by PfamScan against the Pfam database (Mistry et al., 2021). The interacting domains were identified based on the host–pathogen PPIs and collected from the 3 did (three-dimensional interacting domains) database (Mosca et al., 2014). When an interacting domain pair

were present in two proteins, the two proteins were expected to interact with each other.

Performance Evaluation of Predictive Models

To evaluate model performance across different host–pathogen systems, we trained on the dataset in the training host–pathogen systems and evaluated the performance on the test host–pathogen system. In other words, the entire training dataset was partitioned into two parts, i.e., test set and training set. Test set contained dataset from one host and all its pathogens, and training set included all datasets related to the other remaining hosts and pathogens. We used the test set from seven hosts (i.e., *Mus musculus*, *Arabidopsis thaliana*, *Rattus norvegicus*, *Aedes aegypti*, *Bos taurus*, *Sus scrofa*, and *Gallus gallus*) and their pathogens to evaluate the model performance across host–pathogen systems, respectively. For each test host–pathogen system, the process was performed 10 times with the different negative training sets.

The 10-fold cross-validation method was used to evaluate the performance of different models. The training dataset was randomly divided into 10 subsets. Nine of them were combined to train the model, and the remaining one was used to test the model. The progress was repeated 10 times with the different negative training sets, and the final result was the average performance of the 10 replicates. True positive rate (TPR) = TP/(TP + FN), false positive rate (FPR) = FP/(FP + TN), precision = TP/(TP + FP), and F1 score = 2 × (precision × recall)/(precision + recall) were used to evaluate the prediction performance.

Prediction of PPIs Between Rice and Blast Fungus

We built the rice–blast fungus PPI classifier using the random forest from the scikit-learn library in Python. The interaction probability of each protein pair was computed using the optimized model. The protein pair with a probability greater than the threshold of 0.5 was considered to interact with each other.

Identification of Avirulence Genes in Blast Fungal Genome

Gene sequences of avirulence effectors were extracted from the GenBank database (Meng et al., 2019). All predicted coding sequences of the blast fungal genome were then searched against the local avirulence sequence database to identify matches to the cloned genes. The parameters used for the sequence similarity search were ≥95% identity and 80% coverage of the avirulence effector genes.

Analysis of Functional and Pathway Enrichment

Functional enrichment of rice genes was analyzed by using agriGO (Tian et al., 2017) by comparing the reference gene dataset of the rice genome with their False Discovery Rate (FDR) values. The pathway enrichment analysis of rice genes was performed using the Fisher's exact test implemented in a Perl script against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al., 2017).

Network Analysis of Blast Resistance Genes

The main data on blast disease traits were collected from the China Rice Data Center (http://www.ricedata.cn/gene/gene_pi.htm). The blast resistance genes that were identified in the rice reference genome were used in the network analysis.

Transcriptional Analysis of Rice Genes in Response to Blast Fungus

The RNA-seq datasets (accession: SRP079683) derived from rice with *M. oryzae* infection were downloaded from the NCBI SRA database (Bidzinski et al., 2016). The sequencing reads were processed by trimmomatic to remove the adapter sequences and low-quality reads (Bolger et al., 2014). The cleaned reads were mapped to the rice genome using hisat2 (Kim et al., 2019), and the aligned reads were counted with featureCounts (Liao et al., 2014). The differentially expressed genes were identified using DESeq2 with $p < 0.01$ and at least 2-fold changes (Love et al., 2014).

RESULTS

Inference of Host–Pathogen PPIs From Structure Information

We developed a computational method for predicting host–pathogen PPIs based on structural information. The framework is illustrated in **Figure 1**. In brief, given a pair of proteins from host and pathogen, we first predicted protein structures using homology modeling and then searched for their closest PDB complex as a structural template. The interaction model for each protein pair was created by superimposing the homology structures on their corresponding chains in the template complex. Structural features, including structural similarity, structural distance, as well as the number and fraction of the conserved interacting residue pairs, were calculated from the interaction model. Finally, we combined structural evidences and sequence information to predict host–pathogen PPIs using the random forest-based classifiers.

Structure-Based Method Accurately Discovers Plant–Pathogen PPIs

Due to the limited availability of known PPIs between plants and pathogens, cross-species performance is important for the model trained on the dataset in known host–pathogen systems to infer PPIs in a new host–pathogen system. We selected seven test host–pathogen systems to systematically evaluate model performance across host–pathogen systems. For each test, all datasets except from one selected host and all its pathogens were used to train the model, and the dataset from the selected host–pathogen system was used to evaluate the predictive model. As shown in **Figure 2**, the TPR of the method is higher than 68.6% for all test host–pathogen systems. The results indicated its robust
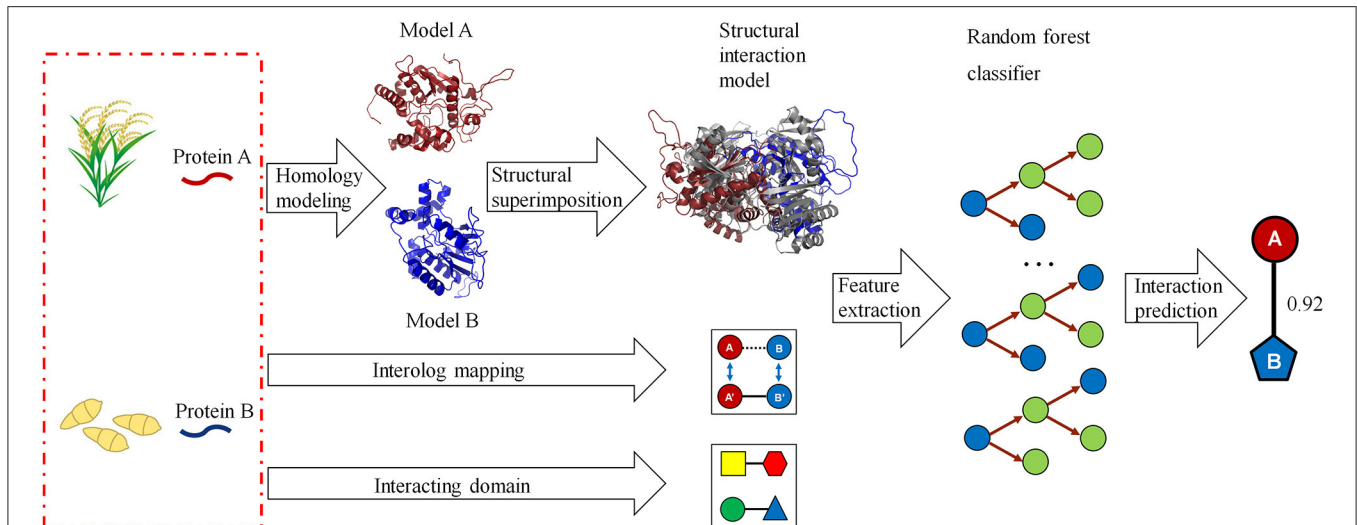
FIGURE 1 | A computational framework for predicting protein–protein interactions (PPIs) between host and pathogen. Given a pair of potentially interacting proteins (A,B) from host and pathogen, three-dimensional structures for each protein pair were built by homology modeling and then searched for their closest PDB complex as a template by structural alignment. Four structure-based scores associated with the protein pair were calculated by superimposing the homology structures on their corresponding chains in the template complex. Finally, combining the structural evidences with the clues of homologous mapping and interacting domains to predict the interaction between proteins A and B using random forest-based classifiers.
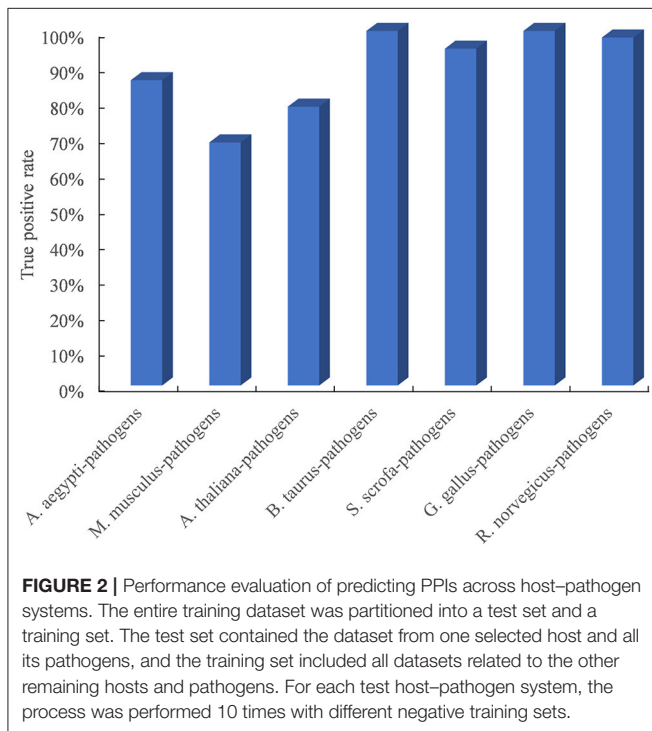
FIGURE 2 | Performance evaluation of predicting PPIs across host–pathogen systems. The entire training dataset was partitioned into a test set and a training set. The test set contained the dataset from one selected host and all its pathogens, and the training set included all datasets related to the other remaining hosts and pathogens. For each test host–pathogen system, the process was performed 10 times with different negative training sets.

TABLE 1 | The performance of different approaches in detecting protein–protein interactions (PPIs) between *Arabidopsis* and pathogen.

Method	Accuracy (ACC)	True positive rate (TPR)	F1-score
Structure-based method	224/256 = 87.5%	101/128 = 78.9%	0.86
domain-based method	148/256 = 57.8%	20/128 = 15.6%	0.27
interolog-based method	151/256 = 59.0%	23/128 = 18.0%	0.30

The test dataset including 128 experimentally determined PPIs and 128 randomly shuffled protein pairs between Arabidopsis and pathogens.

Among the 128 *Arabidopsis*–pathogen PPIs with homology-modeled structures, 101 (78.9%) protein interactions could be successfully predicted by the structure-based method. However, only 23 and 20 *Arabidopsis*–pathogen PPIs were inferred by the homologous mapping and interacting domain pairs, respectively (**Table 1**). For the interaction between *Arabidopsis* and *Ralstonia solanacearum*, 20 of PPIs were detected by the structural similarity, while only 4 of them were supported with the interacting domains (**Supplementary Table 2**). These results indicated that the structure-based method outperformed the interolog and domain-based methods for identifying plant–pathogen PPIs.

Proteome-Wide Prediction of PPIs Between Rice and Blast Fungus

We first trained the prediction model on a dataset consisting of positive and negative examples of equal size. Although the model worked relatively well with an accuracy of 94.13%, the

cross-species prediction performance in discovering PPIs in new host–pathogen systems.

The interolog and domain-based methods are two widely used computational approaches for predicting host–pathogen PPIs. We thus evaluated the performance of our method with respect to these two methods on the test dataset of experimentally determined PPIs between *Arabidopsis* and all its pathogens.

TABLE 2 | Performance comparison of models trained with different ratios of positive and negative samples.

Positive:negative ratio	Accuracy (ACC) (%)	True positive rate (TPR) (%)	False positive rate (FPR) (%)
1:1 (10148: 10148)	94.13	93.63	5.37
1:5 (10148: 50740)	93.09	84.14	5.12
1:10 (10148: 101480)	94.17	66.30	3.04
1:20 (10148: 202960)	96.16	42.79	1.17
1:30 (10148: 304440)	97.30	32.60	0.54
1:50 (10148: 507400)	98.32	23.84	0.19
1:60 (10148: 608880)	98.58	21.75	0.14
1:70 (10148: 710360)	**98.78**	**20.77**	**0.10**
1:80 (10148: 811840)	98.93	19.66	0.08
1:90 (10148: 913320)	99.04	18.90	0.07
1:100 (10148: 1014800)	99.13	18.30	0.06

The bold values mean the optimized positive to negative ratio of training set with a relatively high TPR and a low level of FPR.

FPR of 5.37% would result in a large number of false positive interactions in the proteome-wide prediction of PPIs between rice and blast fungus. Thus, we reduced the FPR by expanding the size of the negative examples in the training dataset. When the ratio of positive to negative samples was adjusted to 1:70 in the training dataset, the FPR was decreased to the expected level of 0.10%, while the TPR remained at the relatively high level of 20.77% (**Table 2**).

To fill the gap between the number of protein sequences and 3D structures, protein structures for rice and blast fungus proteomes were predicted by using homology modeling. These predicted structures contained 32,170 and 2,910 models, covering 82.8 and 83.3% of rice and blast fungus secreted/transmembrane proteomes, respectively. Interaction models for 21,021,571 rice–blast fungus protein pairs were then created by superimposing the homology structures on their corresponding PDB templates.

The optimized host–pathogen prediction model was used to scan all rice–blast fungus protein pairs with the interaction models, resulting in a total of 2,018 PPIs between 1,344 rice proteins and 418 blast fungus proteins (**Figure 3A**, **Supplementary Table 3**). Of the predicted PPIs, only 29.9% (604) of interactions were supported with the evidences of sequence similarity and interacting domain pairs. Moreover, we found that 17 predicted PPIs could be identified by the previous study with the sequence-based method (Ma et al., 2019). These results indicated that the structure-based method could efficiently discover new rice–blast fungus PPIs beyond those interactions inferred from sequence similarity.

As expected, the rice–blast fungus PPI network exhibited scale-free properties similar to those of other biological networks (**Supplementary Figure 1**). It was interesting that the blast fungus proteins had more connections than rice proteins in the PPI network. One blast fungus protein averagely had five interacting partners from rice, while one rice protein interacted with around two blast fungus proteins (**Figure 3B**). Approximately, 12% (50) of blast fungus proteins had at least 10 rice interactors, and the pathogen protein with the highest degree was predicted to interact with 143 rice partners in the network.

The result meant that the potential pathogen-associated proteins had a higher degree than the resistance-associated proteins in the rice–blast fungus PPI network.

Network Captures Key Components in Rice–Blast Fungus Pathosystem

The major components involved in rice–blast fungus interactions include resistance genes from rice and avirulence effectors from *M. oryzae*. Thus, we examined whether the PPI network could predict these key components in rice–blast fungus pathosystem. Currently, about two dozens of blast resistance genes have been cloned and characterized in rice. Among the cloned resistance genes, 12 genes were distributed in the rice reference genome of *O. sativa* spp. *japonica* cv. Nipponbare (**Supplementary Table 4**). Two resistance genes, namely, pi-d2 (*LOC_Os06g29810*) and pi-ta (*LOC_Os12g18360*), were successfully predicted in the PPI network, ∼7-fold enrichment in comparison with that of the whole genome. In addition, a total of 13 avirulence effector genes have been cloned from *M. oryzae*, six of which have the corresponding matches in the reference genome sequence (**Supplementary Table 5**). One avirulence effector, AVR-Pik (MGG_15972), was detected in the PPI network. The AVR-Pik effector was predicted to interact with four resistance-associated proteins (i.e., *LOC_Os02g37290*, *LOC_Os02g37300*, *LOC_Os02g37320*, and *LOC_Os04g39380*) containing a heavy metal-associated domain, which have been validated by experimental measurements (De la Concepcion et al., 2018, 2019). These results indicated the powerful performance of the PPI network in capturing major components in rice–blast fungus interactions.

Functional Analysis of Rice Protein Involved in the Interaction

To determine the function of rice proteins interacted with pathogen proteins, the GO analysis of these proteins in the network was carried out. These resistance-associated proteins were preferentially involved in specific biological processes such as transcriptional regulation, phosphorylation, transmembrane transport, signal transduction, and nucleotide metabolic process (**Figure 4A**). Furthermore, the pathway analysis showed that these rice proteins are significantly enriched in the pathways of spliceosome, ATP-binding cassette (ABC) transporters, plant hormone signal transduction, RNA transport, and oxidative phosphorylation (**Figure 4A**). It was worth to notice that seven rice proteins acted as core components in the pathway of plant–pathogen interactions, including *LOC_Os04g52780*, known as FLS2, an immune receptor, which could activate plant immune response by recognizing flagellin proteins of bacterial pathogens (**Figure 4B**).

Network-Based Prediction of Blast Resistance Gene

The rice blast resistance gene *LOC_Os06g29810* was predicted to interact with pathogen protein *MGG_00990* based on the structural interaction model created by superimposing the homology structures on the template of the AvrPto–Pto complex (**Figure 5A**). The homology model of *MGG_00990* was structurally similar to the pathogen effector AvrPto

FIGURE 3 | Rice–blast fungus PPI network. **(A)** Global view of the rice–blast fungus PPI network. Proteins from rice and blast fungus are represented by blue circles and red triangles, respectively. **(B)** Comparison between the degree distribution of rice and blast fungus proteins in the network. The blue bar represents the degree distribution of rice proteins interacting with blast fungus proteins, and the red bar represents the degree distribution of blast fungus proteins interacting with rice proteins.

FIGURE 4 | Functional and pathway analysis of rice genes in the PPI network. **(A)** Enriched function terms and pathways of rice genes. *Y*-axis represents the GO terms and KEGG pathways, and *X*-axis represents the negative $\log_{10}$(FDR) of enriched terms and pathways. The enriched GO terms and KEGG pathways are represented by blue and red bars, respectively. **(B)** Rice genes in the KEGG pathway of plant–pathogen interaction. The rectangle represents a gene product and the rectangle marked with green color indicates the corresponding rice gene as identified in the rice–blast fungus network.

in the template complex, while the homology model of *LOC_Os06g29810* was structurally close to the plant resistance protein Pto (Xing et al., 2007). The another rice blast resistance gene *LOC_Os12g18360* could interact with the pathogen gene *MGG_08973*, which was inferred from the structural similarity of the homology models to the chains in the structural template of thioredoxin in barley (**Figure 5B**) (Maeda et al., 2008).

The availability of the PPI network allowed the systematic discovery of novel blast resistance genes using the guilt-by-association method. In addition to the two blast resistance genes, 47 rice genes were found to interact with the two

pathogen-associated genes, *MGG_00990* and *MGG_08973*, in the PPI network. The functional analysis revealed that these blast resistance candidates mostly encoded receptor-like cytoplasmic kinases involved in the biological processes of phosphorylation and signaling in rice (**Figure 6A**). Furthermore, the analysis of gene expression showed that these candidates were preferentially responsive to the infection of blast fungus (**Figure 6B**). Among the interacting partners of *MGG_00990*, 18 resistance-associated genes were differentially expressed after blast fungus infection (**Figure 6C**), while 10 genes were significantly induced by the infection blast fungus after drought stress (**Figure 6D**). These

FIGURE 5 | Structural model of protein interaction. **(A)** The structural interaction model for *LOC_Os06g29810* and *MGG_00990* created by superimposing the homology structures on the chains B and A in the template (PDB ID 2qkw). **(B)** The structural interaction model for *LOC_Os12g18360* and *MGG_08973* created by superimposing the homology structures on the chains A and C in the template (PDB ID 2vm1). The homology models of rice and blast fungus proteins are shown in blue and red, respectively. The PDB template complexes are shown in gray.

FIGURE 6 | Functional significance and expression of rice genes in the subnet. **(A)** Enriched GO terms of rice genes in the blast resistance subnet. **(B)** Comparison of differentially expressed genes in rice after blast fungus infection. Rice genes in response to infection by **(C)** blast fungus and **(D)** blast fungus after drought stress in the subnet. The red nodes represent upregulated genes, and the green nodes represent downregulated genes.

results indicated the good performance of the PPI network in discovering rice genes associated with blast resistance.

DISCUSSION

Determining PPIs is an essential step toward understanding the underlying biological mechanism against pathogen infection in plants (Muthamilarasan and Prasad, 2013). In this study, we presented a computational method for predicting plant–pathogen PPIs based on structural information. Using seven host–pathogen systems, we demonstrated that the structure-based method was powerful in discovering protein interactions across different host–pathogen systems. This advantage was of vital importance for machine learning-based method for predicting PPIs in new plant–pathogen systems based on prior knowledge obtained from known interacting proteins in other host–pathogen systems.

Although many experimentally validated host–pathogen PPIs have been deposited in the database, most of these PPIs (42,972 out of 45,200) focus on the protein interactions between humans and pathogens (Ammari et al., 2016). Due to the limited availability of plant–pathogen PPIs, many plant-specific pathogen effector proteins usually failed to identify any homolog in the known inter-species interactions using homologous sequence mapping (Yang et al., 2019). In addition, the intra-species PPIs from model organisms have been usually used to infer plant–pathogen PPIs (Li et al., 2012; Sahu et al., 2014). However, evolutionary differences between inter-species and intra-species PPI interfaces would limit the performance of plant–pathogen PPIs by transferring interactions across species (Franzosa and Xia, 2011). These weaknesses could be partially overcome by utilizing structural information. The developed method uses the structural similarity between proteins as a bridge to identify new interactions across plant–pathogen systems. Compared with the sequence-based methods, this structure-based approach enabled us to discover new interactions between plant and pathogen proteins that lacked significant sequence similarity with a known interaction template.

Magnaporthe oryzae is a notorious plant pathogen that causes the most destructive diseases of rice in the world. The prediction and analysis of PPIs are valuable in deciphering the molecular mechanisms of rice–blast fungus interactions. In this study, we generated a global rice–blast fungus PPI network that consisted of 2,018 interacting protein pairs involving 1,344 rice proteins and 418 blast fungus proteins. Over 70% of PPIs between rice and blast fungus were inferred from structural similarity, which greatly expanded the landscape of the rice–blast fungus PPI network. Compared with the previous PPI network, 17 of PPIs were identified by the structure-based method (Ma et al., 2019). Although the number of common PPIs was relatively small, the results were significantly overlapped between the two independent studies (i.e., Fisher's exact test $p < 5.6e-39$). Moreover, we noted that blast fungus proteins had more interacting partners than rice proteins in the network. Our findings were coherent with the sequence-based studies in which

a few pathogen-associated proteins were involved in the plant–pathogen interactions (Li et al., 2012; Sahu et al., 2014; Ma et al., 2019). This is likely to be the result of the coevolutionary arms races, in which pathogens mutate genes extensively to infect their hosts, while plants defend against pathogen attacks by expanding gene families (Stahl and Bishop, 2000; Dangl and McDowell, 2006).

Despite the advances made in molecular mechanisms of rice resistance to blast fungus, many aspects of the rice immunity system remain obscure (Li et al., 2019). The rice–blast fungus PPI network showed that the AVR-Pik effector was successfully predicted to interact with four rice proteins, which have been validated by experimental approaches (De la Concepcion et al., 2018, 2019). In addition to the avirulence effector, two rice blast resistance genes were also identified in the network. Using the guilt-by-association method, we identified 47 candidate blast resistance genes in the PPI network. The majority of these genes that encoded receptor-like cytoplasmic kinases were involved in the response to the infection of blast fungus (Bidzinski et al., 2016). The PPI network provided a global map to help accelerate the identification of blast resistance genes and advance our understanding of the molecular mechanisms of plant–pathogen interactions.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.690124/full#supplementary-material

Supplementary Figure 1 | Degree distribution of the node proteins.

Supplementary Table 1 | Availability of protein–protein interactions (PPIs) in model organisms from various databases.

Supplementary Table 2 | The experimentally determined PPIs between *Arabidopsis* and pathogens.

Supplementary Table 3 | The predicted PPIs between rice and blast fungus.

Supplementary Table 4 | The isolated dominant blast resistance genes in rice reference genome.

Supplementary Table 5 | The isolated avirulence genes in *Magnaporthe oryzae* genome.

REFERENCES

Alfarano, C., Andrade, C. E., Anthony, K., Bahroos, N., Bajec, M., Bantoft, K., et al. (2005). The biomolecular interaction network database and related tools 2005 update. *Nucleic Acids Res.* 33, D418–D424. doi: 10.1093/nar/gki051

Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., et al. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37, 420–423. doi: 10.1038/s41587-019-0036-z

Ammari, M. G., Gresham, C. R., McCarthy, F. M., and Nanduri, B. (2016). HPIDB 2.0: a curated database for host-pathogen interactions. *Database* 2016:baw103. doi: 10.1093/database/baw103

Bidzinski, P., Ballini, E., Ducasse, A., Michel, C., Zuluaga, P., Genga, A., et al. (2016). Transcriptional basis of drought-induced susceptibility to the rice blast fungus *Magnaporthe oryzae*. *Front. Plant Sci.* 7:1558. doi: 10.3389/fpls.2016.01558

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30:2114–2120. doi: 10.1093/bioinformatics/btu170

Burley, S. K., Berman, H. M., Kleywegt, G. J., Markley, J. L., Nakamura, H., and Velankar, S. (2017). Protein Data Bank (PDB): the single global macromolecular structure archive. *Methods Mol. Biol.* 1607, 627–641. doi: 10.1007/978-1-4939-7000-1_26

Calderone, A., Iannuccelli, M., Peluso, D., and Licata, L. (2020). Using the MINT database to search protein interactions. *Curr. Protoc. Bioinformatics*. 69:e93. doi: 10.1002/cpbi.93

Cao, F. Y., Khan, M., Taniguchi, M., Mirmiran, A., Moeder, W., Lumba, S., et al. (2019). A host-pathogen interactome uncovers phytopathogenic strategies to manipulate plant ABA responses. *Plant J.* 100, 187–198. doi: 10.1111/tpj.14425

Cesari, S., Thilliez, G., Ribot, C., Chalvon, V., Michel, C., Jauneau, A., et al. (2013). The rice resistance protein pair RGA4/RGA5 recognizes the *Magnaporthe oryzae* effectors AVR-Pia and AVR1-CO39 by direct binding. *Plant Cell*. 25, 1463–1481. doi: 10.1105/tpc.112.107201

Dangl, J. L., and McDowell, J. M. (2006). Two modes of pathogen recognition by plants. *Proc. Natl. Acad. Sci. U.S. A.* 103:8575–8576. doi: 10.1073/pnas.0603183103

Davis, F. P., and Sali, A. (2005). PIBASE: a comprehensive database of structurally defined protein interfaces. *Bioinformatics* 21, 1901–1907. doi: 10.1093/bioinformatics/bti277

De la Concepcion, J. C., Franceschetti, M., MacLean, D., Terauchi, R., Kamoun, S., Banfield, M. J. (2019). Protein engineering expands the effector recognition profile of a rice NLR immune receptor. *Elife* 8:e47713. doi: 10.7554/eLife.47713

De la Concepcion, J. C., Franceschetti, M., Maqbool, A., Saitoh, H., Terauchi, R., Kamoun, S., et al. (2018). Polymorphic residues in rice NLRs expand binding and response to effectors of the blast pathogen. *Nat. Plants* 4, 576–585. doi: 10.1038/s41477-018-0194-x

Dean, R. A., Talbot, N. J., Ebbole, D. J., Farman, M. L., Mitchell, T. K., Orbach, M. J., et al. (2005). The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* 434, 980–986. doi: 10.1038/nature03449

Deng, M., Mehta, S., Sun, F., and Chen, T. (2002). Inferring domain-domain interactions from protein-protein interactions. *Genome Res.* 12, 1540–1548. doi: 10.1101/gr.153002

Franzosa, E. A., and Xia, Y. (2011). Structural principles within the human-virus protein-protein interaction network. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10538–10543. doi: 10.1073/pnas.1101440108

Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 35, W585–W587. doi: 10.1093/nar/gkm259

International Rice Genome Sequencing Project. (2005). The map-based sequence of the rice genome. *Nature* 436, 793–800. doi: 10.1038/nature03895

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361. doi: 10.1093/nar/gkw1092

Kanzaki, H., Yoshida, K., Saitoh, H., Fujisaki, K., Hirabuchi, A., Alaux, L., et al. (2012). Arms race co-evolution of *Magnaporthe oryzae* AVR-Pik and rice Pik genes driven by their physical interactions. *Plant J.* 72, 894–907. doi: 10.1111/j.1365-313X.2012.05110.x

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. doi: 10.1038/s41587-019-0201-4

Li, W., Chern, M., Yin, J., Wang, J., and Chen, X. (2019). Recent advances in broad-spectrum resistance to the rice blast disease. *Curr. Opin. Plant Biol.* 50, 114–120. doi: 10.1016/j.pbi.2019.03.015

Li, Z. G., He, F., Zhang, Z., and Peng, Y. L. (2012). Prediction of protein-protein interactions between *Ralstonia solanacearum* and *Arabidopsis thaliana*. *Amino Acids*. 42, 2363–2371. doi: 10.1007/s00726-011-0978-z

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Liu, S., Liu, Y., Zhao, J., Cai, S., Qian, H., Zuo, K., et al. (2017). A computational interactome for prioritizing genes associated with complex agronomic traits in rice (Oryza sativa). *Plant J.* 90, 177–188. doi: 10.1111/tpj.13475

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Ma, S., Song, Q., Tao, H., Harrison, A., Wang, S., Liu, W., et al. (2019). Prediction of protein-protein interactions between fungus (*Magnaporthe grisea*) and rice (Oryza sativa L.). *Brief Bioinform*. 20, 448–456. doi: 10.1093/bib/bbx132

Maeda, K., Hägglund, P., Finnie, C., Svensson, B., and Henriksen, A. (2008). Crystal structures of barley thioredoxin h isoforms HvTrxh1 and HvTrxh2 reveal features involved in protein recognition and possibly in discriminating the isoform specificity. *Protein Sci.* 17, 1015–1024. doi: 10.1110/ps.083460308

Marcotte, E. M., Pellegrini, M., Ng, H. L., Rice, D. W., Yeates, T. O., and Eisenberg, D. (1999). Detecting protein function and protein-protein interactions from genome sequences. *Science* 285, 751–753. doi: 10.1126/science.285.5428.751

Mariano, R., and Wuchty, S. (2017). Structure-based prediction of host-pathogen protein interactions. *Curr. Opin. Struct. Biol.* 44, 119–124. doi: 10.1016/j.sbi.2017.02.007

Matthews, L. R., Vaglio, P., Reboul, J., Ge, H., Davis, B. P., Garrels, J., et al. (2001). Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs". *Genome Res.* 11, 2120–2126. doi: 10.1101/gr.205301

Meng, Q., Gupta, R., Min, C. W., Kwon, S. W., Wang, Y., Je, B. I., et al. (2019). Proteomics of rice-*Magnaporthe oryzae* interaction: what have we learned so far? *Front. Plant Sci.* 10:1383. doi: 10.3389/fpls.2019.01383

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2021). Pfam: the protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi: 10.1093/nar/gkaa913

Möller, S., Croning, M. D., and Apweiler, R. (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* 17, 646–653. doi: 10.1093/bioinformatics/17.7.646

Mosca, R., Céol, A., Stein, A., Olivella, R., and Aloy, P. (2014). 3did: a catalog of domain-based interactions of known three-dimensional structure. *Nucleic Acids Res.* 42, D374–D379. doi: 10.1093/nar/gkt887

Mukhtar, M. S., Carvunis, A. R., Dreze, M., Epple, P., Steinbrenner, J., Moore, J., et al. (2011). Independently evolved virulence effectors converge onto hubs in a plant immune system network. *Science* 333, 596–601. doi: 10.1126/science.1203659

Muthamilarasan, M., and Prasad, M. (2013). Plant innate immunity: an updated insight into defense mechanism. *J. Biosci.* 38, 433–449. doi: 10.1007/s12038-013-9302-2

Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., et al. (2014). The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res*. 42, D358–D363. doi: 10.1093/nar/gkt1115

Oughtred, R., Rust, J., Chang, C., Breitkreutz, B. J., Stark, C., Willems, A., et al. (2021). The BioGRID database: a comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci*. 30, 187–200. doi: 10.1002/pro.3978

Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D., and Yeates, T. O. (1999). Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl. Acad. Sci. U.S.A.* 96, 4285–4288. doi: 10.1073/pnas.96.8.4285

Pieper, U., Webb, B. M., Dong, G. Q., Schneidman-Duhovny, D., Fan, H., Kim, S. J., et al. (2014). ModBase, a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res*. 42, D336–D346. doi: 10.1093/nar/gkt1144

Sahu, S. S., Weirick, T., and Kaundal, R. (2014). Predicting genome-scale Arabidopsis-Pseudomonas syringae interactome using domain and interolog-based approaches. *BMC Bioinform*. 11:S13. doi: 10.1186/1471-2105-15-S 11-S13

Salwinski, L., Miller, C. S., Smith, A. J., Pettit, F. K., Bowie, J. U., and Eisenberg, D. (2004). The database of interacting proteins: 2004 update. *Nucleic Acids Res*. 32, D449–D451. doi: 10.1093/nar/gkh086

Stahl, E. A., and Bishop, J. G. (2000). Plant-pathogen arms races at the molecular level. *Curr. Opin. Plant Biol*. 3, 299–304. doi: 10.1016/S1369-5266(00)00083-2

Tanwar, H., and George Priya Doss, C. (2018). Computational resources for predicting protein-protein interactions. *Adv. Protein Chem. Struct. Biol*. 110, 251–275. doi: 10.1016/bs.apcsb.2017.07.006

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., et al. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res*. 45, W122–W129. doi: 10.1093/nar/g kx382

Weßling, R., Epple, P., Altmann, S., He, Y., Yang, L., Henz, S. R., et al. (2014). Convergent targeting of a common host protein-network by pathogen effectors from three kingdoms of life. *Cell Host Microbe*. 16, 364–375. doi: 10.1016/j.chom.2014. 08.004

Xing, W., Zou, Y., Liu, Q., Liu, J., Luo, X., Huang, Q., et al. (2007). The structural basis for activation of plant immunity by bacterial effector protein AvrPto. *Nature* 449, 243–247. doi: 10.1038/nature06109

Yang, S., Li, H., He, H., Zhou, Y., and Zhang, Z. (2019). Critical assessment and performance improvement of plant-pathogen protein-protein interaction prediction methods. *Brief Bioinform*. 20, 274–287. doi: 10.1093/bib/bbx123

Zhang, F., Liu, S., Li, L., Zuo, K., Zhao, L., and Zhang, L. (2016). Genome-wide inference of protein-protein interaction networks identifies crosstalk in abscisic acid signaling. *Plant Physiol*. 171, 1511–1522. doi: 10.1104/pp.16.00057

Zhang, Q. C., Petrey, D., Deng, L., Qiang, L., Shi, Y., Thu, C. A., et al. (2012). Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature* 490, 556–560. doi: 10.1038/nature11503

Zhang, Y., and Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res*. 33, 2302–2309. doi: 10.1093/nar/gki524

Zhao, J., Lei, Y., Hong, J., Zheng, C., and Zhang, L. (2019). AraPPINet: An Updated Interactome for the Analysis of Hormone Signaling Crosstalk in *Arabidopsis thaliana. Front. Plant Sci*. 10:870. doi: 10.3389/fpls.2019.00870

Comprehensive *in silico* Characterization of Universal Stress Proteins in Rice (*Oryza sativa* L.) With Insight Into Their Stress-Specific Transcriptional Modulation

Shatil Arabia[†‡], Asif Ahmed Sami[†‡], Saima Akhter[†], Rakha Hari Sarker[†] and Tahmina Islam*[†]

Plant Breeding and Biotechnology Laboratory, Department of Botany, University of Dhaka, Dhaka, Bangladesh

In a world where climate change is real and its consequences are unprecedented, understanding of the plant adaptive capacity and native stress-responsive machinery is crucial. In recent years, universal stress proteins (USPs) have received much attention in the field of plant science due to their stress-specific transcriptional regulation. This study focuses on the extensive characterization of the *USP* gene family members in the monocot crop rice (*Oryza sativa* L. var. *japonica*). Here, we report a total of 44 *USP* genes in the rice genome. *In silico* characterization of these genes showed that domain architecture played a major role in the functional diversification of the *USP* gene family which holds for all plant *USPs*. On top of that, a higher conservation of *OsUSP* members has been exhibited with a monocot genome (*Zea mays* L.) as compared to a dicot genome (*Arabidopsis thaliana* L.). Expression profiling of the identified genes led to the discovery of multiple *OsUSP* genes that showed pronounced transcript alteration under various abiotic stress conditions, indicating their potential role as multi-functional stress-specific modules. Furthermore, expression validation of *OsUSP* genes using qRT-PCR provided a strong evidence for the utility *OsUSP* genes in building multi-stress tolerant plants. Altogether, this study provides leads to suitable *USP* candidates that could be targeted for plant breeding and genetic engineering experiments to develop stress resilient crop species.

Keywords: universal stress proteins, rice, transcript alteration, abiotic stress, functional validation

INTRODUCTION

Alteration of the growth environment from optimum to adverse conditions is a common occurrence in a plant's life cycle. Being sessile, plants must exhibit a dynamic response to cope with subtle to drastic changes in their nearby regimes. Such changes can range from diverse ecological parameters such as temperature, water and nutrient availability, salt content, etc. to intrinsic factors such as accumulation of reactive oxygen species (ROS). Under such unfavorable conditions, plants can suffer numerous repercussions, e.g., arrested growth and development,

lower photosynthetic capacity, abnormalities in flowering time, reduced fertility and germination rate, abated total yield, etc. (Munns, 2002; Barnabás et al., 2008). To overcome this, plants have evolved highly complex yet coordinated responses through temporal and spatial regulation of genes that can mitigate unwanted effects resulting from various abiotic stressors. Generally, perception of stressors triggers many downstream signaling cascades within plant cells, such as protein kinases, phosphatases, transcription factors, molecular chaperones and defense-related proteins that mediate a suitable response to ensure plant survival (Jung et al., 2015). However, in a world where climate change is experiencing a rapid spike, plants native stress-responsive machinery is not able to protect the cellular system from stress inflicted damages (Lesk et al., 2016). These circumstances have accelerated the search for stress-induced genetic components that hold the capacity to endow plants with enhanced stress resilience through modern biotechnological tools. In this regard, the most suitable candidates include those genes that can mediate protection against multiple stressors simultaneously. One such group of proteins that have recently come into the spotlight for their stress-specific roles in plants, are the universal stress protein (USP) family (Chi et al., 2019).

The first USP was reported in *E. coli* where the candidate protein responded to a multitude of environmental perturbations including nutrient starvation, temperature shock, oxidative and osmotic stress, heavy metal toxicity, antibiotics, etc. (Nyström and Neidhardt, 1992; Zarembinski et al., 1998; Sousa and McKay, 2001). Later, USPs have been reported in many other organisms including other bacteria, archaea, fungi, and even plants (Vollmer and Bark, 2018). Although *E. coli* USPs by structure only contain the USP-domain, plant USPs have largely diversified in function by accumulating additional accessory domains that allow them to participate in numerous varying functions (Chi et al., 2019). Consequently, the multi-functional nature of USPs is derived from their structural diversity. Information of the evolution and function of plant USPs is very limited to date. A few genes have been studied in rice, *Arabidopsis,* cotton, etc. which were found to play a key role in stress regulation (Sauter, 2002; Merkouropoulos and Tsaftaris, 2013; Jung et al., 2015). Nevertheless, these studies strongly hint to the importance of plant USPs for further studies. *OsUSP1* from rice was the first *USP* gene identified in the plant kingdom that has been involved in the activation of signaling cascade in response to ethylene during hypoxia (Sauter, 2002). It was reported that stomatal closure is enhanced by *SpUSP* during drought stress to limit the detriment in tomatoes (Loukehaich et al., 2012). Two *Arabidopsis thaliana USP* isoforms (At3g62550 and At3g53990) were reported to show enhanced expression under drought (Shinozaki and Yamaguchi-Shinozaki, 2007). Recent studies have also highlighted the role of USPs in conferring benefits during other abiotic stresses (Udawat et al., 2016). For instance, the *AtUSP* (At3g53390) gene is significantly induced by salt, osmotic stress, and wounding (Isokpehi et al., 2011). Overexpressing *AtUSP* (At3g53990) conferred tolerance against oxidative and heat stresses whereas sensitivity to these stressors was observed in *atusp* (SALK_146059) mutant lines (Jung et al., 2015). *AtUSP* was also found to be involved in cold stress (Melencion et al., 2017). Apart from that, *USP*

regulates ethylene-mediated signaling and thereby modulates fruit ripening (Isokpehi et al., 2011). A tomato USP protein, SlRd2 interacts with calcineurin B-like interacting protein kinase (SlCipk6) and is involved in the regulation of ROS production (Gutiérrez-Beltrán et al., 2017). In cotton, two closely related genes *GhUSP1* and *GhUSP2*, within the USP family were reported to be drought-responsive (Maqbool et al., 2009). Moreover, *AtUSP* promoter has shown to upregulate *GUS* expression in response to abscisic acid (ABA), 1-aminocyclopropane-1-carboxylic acid (ACC), dehydration, heat, cold, salt, and osmotic stress (Bhuria et al., 2016). This is a clear indication of the functional capacity of plant USPs under multi-stress conditions. In line with such importance, forty-one *USP* genes were identified in the model plant *Arabidopsis thaliana* (Bhuria et al., 2019). Out of these, *AtUSP9* and *AtUSP12* were identified as multiple stress-responsive.

Rice is an important crop and an ideal model for the comparative study of gene functions owing to high efficiency of genetic transformation and ease of cultivation and propagation (Basso et al., 2020). With the release of the whole genome sequence of rice (Song et al., 2019), it is very convenient to comprehensively analyze an entire gene family. Recently, 38 *USP* genes were reported in rice (Japonica variety) (Chi et al., 2019). However, a detailed analysis of the functional annotation and mode of regulation of these genes is still lacking. Due to the multi-stress responsive nature of plant USPs, it is imperative to look for the suitable candidates that could be targeted for improving crop resilience under multiple environmental adversities. Here, considering the important role of USP proteins in plant stress regulation and the lack of information about this gene family in a crop species, our study aimed to screen the *USP* gene family in rice. This study provides an in-depth bioinformatics characterization of the identified *USP* genes along with their developmental and stress-specific transcriptional regulation. Overall, this study provides suitable *USP* genes for future biotechnological applications aiming to generate of multi-stress tolerant agricultural crops and serves as a framework for investigating the *USP* gene family in other important plant species.

MATERIALS AND METHODS

Identification, Gene Nomenclature and *in silico* Analysis of *USP* Genes in Rice

To identify the *USP* genes in rice, blastp and Hidden Markov Model (HMM) search had been done. Rice [*Oryza sativa Japonica* Group (Japanese rice) cv. Nipponbare] USP protein sequences were retrieved from the publicly available genome database in the Rice Annotation Project Database (RAP-DB[1]) (Sakai et al., 2013) using *Arabidopsis USP* (At1g68300) as a query sequence. Subsequently, each of the newly identified sequences were used as a query sequence to find the other new members. To confirm the result from HMM and blastp search, the putative USP sequences

---

[1] http://rapdb.dna.affrc.go.jp/

were also analyzed against the Pfam database[2] (Mistry et al., 2020). Nomenclature of rice *USPs* was given according to the chromosomal location in descending order (Hasan et al., 2021). Information on the gene locus ID, protein size and full-length cDNA, the molecular weight, theoretical pI, the number of amino acids were obtained from the RAP-DB database. The subcellular localization of the proteins was predicted using CELLO v.2.5: subCELlular LOcalization predictor[3] (Yu et al., 2006) and WoLF pSORT[4] software (Horton et al., 2007). ChloroP[5] (Emanuelsson et al., 1999) was used to verify chloroplast localization.

Exon–Intron Distribution of *OsUSP* Genes, Identification of Conserved Motifs and Amino Acid Content in OsUSP Proteins

The GFF3 annotation file for the rice genome was downloaded from the Ensembl Plants[6] (Howe et al., 2020). Information regarding the position of exon–introns and UTR regions of rice *USP* genes were extracted from this file. For the identification of conserved motifs, MEME SUITE software (v.5.3.3) was used (Bailey et al., 2009) with all parameters were kept in default settings except the number of motifs to find was set to 10. The "Gene Structure View" option of TBtools (v.1.092) (Chen et al., 2020) was used to visualize the exon–intron distribution and conserved motifs of rice USP members. Both the modified GFF3 file and MEME SUITE XML output file was used for this task. Amino acid content was calculated using the Biopython (v.1.78) package.

Chromosomal Localization of *OsUSP* Genes and Their Synteny Analysis With *Arabidopsis* and Maize Genome

Chromosomal distribution of each *OsUSP* gene was retrieved from the Oryzabase database for rice[7]. For the duplication study within the rice genome, data including the synonymous rate (Ks) and non-synonymous rate (Ka) values were retrieved from the plant genome duplication database[8] (Lee et al., 2012). Duplication type and selection pressure was calculated according to Hasan et al. (2021). The estimated date (Mya, million years ago) of each duplication event was calculated by using $T = Ks/2\lambda$ where T is divergence time, Ks is the number of synonymous substitutions per site, and λ is the fixed substitution rate of 6.5×10^{-9} mutations per synonymous site per year in grasses (Gaut et al., 1996). Syntenic relationships between rice, *Arabidopsis* and maize *USP* genes were constructed and visualized using MCScanX (Wang et al., 2012) and TBtools (v.1.092) (Chen et al., 2020), respectively. In TBtools, the "Advanced Circos" option was used for visualizing the syntenic blocks.

Phylogenetic Relationship Between Rice, *Arabidopsis*, and *E. coli* USP Proteins Along With Their Domain Architecture

To get an idea about the evolutionary history of the rice *USP* gene family, we performed a phylogenetic analysis using the USP protein sequences of rice, *Arabidopsis*, and *E. coli*. *Arabidopsis* and *E. coli* USP protein sequences were retrieved from TAIR (Berardini et al., 2015) and UniProt (UniProt Consortium, 2021) databases, respectively. The protein sequences were aligned using the online version of MAFFT tool v.7 (Katoh et al., 2019) with default parameters. Next, gaps in the aligned regions were trimmed using the Phyutility tool (v.2.7.1) (Smith and Dunn, 2008) to retain 75% of the consensus alignment. With the trimmed multiple sequence alignment file, a phylogenetic tree was constructed using IQ-Tree v.2.1.2 (Minh et al., 2020) with 1000 ultrafast bootstrap replicates. Finally, the tree was visualized and edited using iTOL (v.6) (Letunic and Bork, 2019). The domain features were uploaded as an additional dataset in the required format (details on the iTOL help page). Domain information of each USP protein was downloaded from Pfam (Finn et al., 2014).

Identification of Microsatellite Markers, Glycosylation and Phosphorylation Sites in OsUSPs

For the identification of SSR markers in *OsUSP* members, a microsatellite identification tool (MISA[9]) (Beier et al., 2017) was used. The minimum repeat unit was defined as 10 units for mononucleotide repeats and 5 units for dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide repeats. The maximum length of sequence between two markers was set to 100 units. The number of glycosylation sites in OsUSP proteins was predicted using NetNGlyc 1.0 server[10] (Gupta and Brunak, 2002) with a threshold of 0.5. The predicted phosphorylation sites of all OsUSP members were identified using NetPhos3.1[11] (Blom et al., 1999, 2004) with the threshold of 0.75. Predictions were performed with all three residues (Tyr, Thr, Ser) for both generic and kinase-specific phosphorylation.

Expression Analysis of *OsUSP* Genes at Different Tissues, Developmental Stages and Under Stress Conditions Using Publicly Available mRNA-Seq Data

Temporal and spatial expression profiles of *OsUSP* genes were retrieved from Genevestigator[12] (Hruz et al., 2008) for different developmental stages, anatomical tissues and in response to stress conditions (**Supplementary Tables 1–3**). Heatmaps were generated for each type of expression data using the "Heatmap Illustrator" option in TBtools v.1.092 (Chen et al., 2020).

---

[2] http://pfam.sanger.ac.uk/

[3] http://cello.life.nctu.edu.tw/

[4] https://www.genscript.com/wolf-psort.html

[5] http://www.cbs.dtu.dk/services/ChloroP/

[6] https://plants.ensembl.org/info/data/ftp/index.html

[7] http://viewer.shigen.info/oryzavw/maptool/MapTool.do

[8] http://chibba.agtec.uga.edu/duplication/index/downloads

[9] http://pgrc.ipk-gatersleben.de/misa/misa.html

[10] http://www.cbs.dtu.dk/services/NetNGlyc/

[11] https://services.healthtech.dtu.dk/service.php?NetPhos-3.1

[12] https://genevestigator.com/gv/index.jsp

Heatmaps of both tissue and developmental expression patterns were created using the log_{10} transformed mean expression values. As for the expression under stress conditions, the log_2 fold change in expression data was used to generate the Heatmap.

Analysis of *Cis*-Regulatory Elements and GO Enrichment of *OsUSP* Genes

To identify the *cis*-regulatory elements present in the putative promoter region, 1 kb upstream sequence of each *OsUSP* gene was used. Subsequently, the sequences were uploaded to the PlantCARE database[13] (Lescot, 2002) to detect the presence of various *cis*-elements. Later, the presence of these *cis*-regulatory elements was visualized using a stacked bar plot. Gene Ontology enrichment analysis of *OsUSP* genes was performed with the help of agriGO database v.2 (Tian et al., 2017). The results were visualized in the R programming platform.

Plant Growth and Stress Treatments

For expression profiling, a uniformly developed rice seeds of BRRI-53 variety were placed in a greenhouse (16 h photoperiod, $28 \pm 2°C$ temperature) (Islam et al., 2015). The 15 days old seedlings were then taken for various experimental treatment such as salt (200 mM), drought (mannitol; 150 mM), oxidative (30% H_2O_2), cold (4°C), heat (42°C), and dehydration (air-dried), ABA (10 mM), gibberellic acid (GA3, 1 mM), salicylic acid (SA, 2 mM) (Islam et al., 2015). The untreated seedlings were considered as a control to the treated seedlings. After 16 h treatment, leaves were collected (three biological replicates), immediately frozen in liquid nitrogen, and stored at $-80°C$.

RNA Isolation and RT-PCR Analysis

Rice seeds of BRRI-53 variety were grown in the controlled environmental condition of $28 \pm 2°C$ temperature and 16 h photoperiod in a growth chamber. Fifteen days' old seedlings were then used for various abiotic stress treatment such as salt (200 mM NaCl), drought (150 mM Mannitol), oxidative (30% H_2O_2), cold (4°C), heat (42°C), dehydration (air-dried) and hormonal treatment such as, ABA, GA3, and SA (Islam et al., 2015), for 16 h. The untreated seedlings were considered as a control to all these stresses. The tissues were harvested, and total plant RNA was isolated from rice shoots using TRIzol reagent (Invitrogen, United States) as per the manufacturer's protocol (**Supplementary Figure 1**).

First-strand cDNA was synthesized with reverse transcriptase (Thermo Fisher Scientific, United States) by following the manufacturer's protocol. A reaction mix (12 μl) was prepared in a PCR tube containing 10 μl of DNase treated RNA, 1 μl of DEPC-H_2O and 1 μl oligo dT primer. The reaction mixture was incubated at 65°C for 5 min and chilled down on ice immediately for 5 min. Then, a master mix of 4 μl containing 5X reaction buffer, 1 μl of Riboblock RNAse inhibitor, 1 μl of Revertaid and 2 μl of 10 mM dNTPs mix was added in the tubes. The mixture was incubated at 42°C for 60 min. After the first strand synthesis, the reaction was terminated by heat

inactivation at 70°C for 5 min. The expression of individual genes was measured with gene-specific primers by real-time PCR analysis with a cycler Applied biosystem 7500 and SYBR Green mixture (Bio-Rad, United States) (**Supplementary Table 4**). The relative expression of specific genes was quantitated with the 2-$\Delta\Delta Ct$ calculation method (Livak and Schmittgen, 2001), where $\Delta\Delta Ct$ was the difference in the threshold cycles and the reference gene, which was rice *eEF1* for the expression analyses (Islam et al., 2015). The sequences of gene-specific primers are provided in **Supplementary Table 5**.

Statistical Analysis

Statistical analysis was performed for the relative normalized expression data from three biological replicates under each treatment ($n = 3$). The paired student's *t*-test was performed for each treatment against the respective controls to determine the significance level that were marked with *, **, *** in case of *P*-value < 0.05, <0.01, and <0.001; respectively.

Homology Based Modeling and Molecular Docking Study

Three-dimensional structures of nine USPs were created using the MODELLER tool of MPI bioinformatics toolkit[14] (Zimmermann et al., 2018; Gabler et al., 2020) based on the templates identified using the HHpred tool. Crystal Structure of the Usp protein of *Mycobacteria* (PDB-5AHW_F), hypothetical protein MJ0577 (PDB-1MJH_B), UspA from *Lactobacillus plantarum* (PDB-3S3F_A), USP from *A. thaliana* (PDB-3GM3_E), USP from *Nitrosomonas* (PDB-3TNJ_A), USP from *Burkholderia pseudomallei* (PDB-4WNY_A), Human Cdkl5 Kinase Domain (PDB-4BGQ_A), Pseudokinase MLKL from *Mus musculus* (PDB-4BTF_A) and hypothetical protein PH0823 (PDB-2DUM_B) were used as templates for OsUSP2, OsUSP3, OsUSP6, OsUSP14, OsUSP22, OsUSP32, OsUSP33 and OsUSP36, respectively. All these structures were visualized using BIOVIA Discovery Studio v.21.1.0.20298 (BIOVIA, 2021) (**Supplementary Figure 2**).

3D structure of the above nine OsUSP proteins was used as a receptor to find out the binding affinity with three well-established inhibitors of *E. coli* UspA – ZINC000104153710, ZINC000004268284, ZINC000000217308 (Bandyopadhyay et al., 2021). Four kinase inhibitors – fisetin, luteolin, myricetin, quercetin (Cassidy and Setzer, 2010) were used to check their binding affinity toward the kinase domain-containing OsUSPs. The 3-D chemical structure of these ligands was downloaded from the PubChem compound database[15] (Kim et al., 2021) as an SDF file. The overall docking process was done using PyRx Virtual Screening software (Dallakyan and Olson, 2015) which uses Open Babel (O'Boyle et al., 2011) for importing SDF files, removing salts and energy minimization, uses AutoDock tools (Morris et al., 2009) for the preparation of protein and to generate input files and AutoDock Vina (Trott and Olson, 2010) as docking wizard. Grid box parameters were set to perform a blind docking (Hetényi and van der Spoel, 2002)

---

[13]http://bioinformatics.psb.ugent.be/webtools/plantcare/html/

[14]http://toolkit.tuebingen.mpg.de
[15]http://www.pubchem.ncbi.nlm.nih.gov
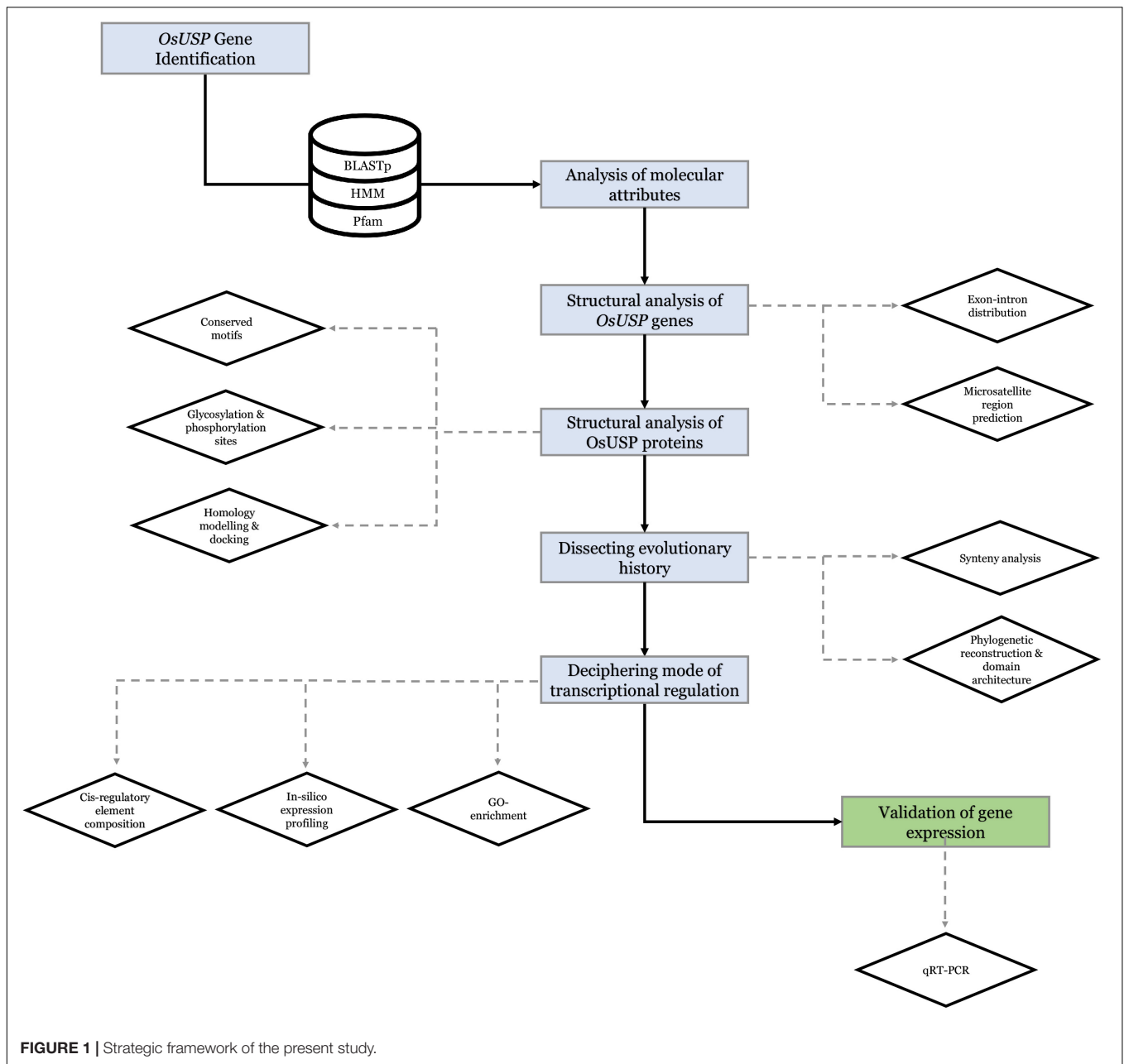
FIGURE 1 | Strategic framework of the present study.

of the inhibitors on the targeted proteins. The 2-D and 3-D interaction of the protein–ligands complex has been observed by using the BIOVIA Discovery Studio v.21.1.0.20298 (BIOVIA, 2021). Overall, a strategic framework of the present study has been summarized in **Figure 1**.

RESULTS

Identification of *USP* Genes in Rice

Based on the blastp and HMM search, a total of 44 candidate *USP* genes were identified in rice (*Oryza sativa* L.). Relevant details about these genes are presented in **Table 1**. These

identified *USP* genes were named from *OsUSP1* to *OsUSP44* in order of their chromosomal position. Among them, two splice variants were observed for genes *OsUSP1*, *OsUSP2*, *OsUSP11*, *OsUSP16*, *OsUSP20*, *OsUSP22*, *OsUSP25*, *OsUSP27*, *OsUSP30*, *OsUSP31*, and *OsUSP35*; whereas, *OsUSP12* had three splice variants. The length of deduced complementary DNA sequence (CDS) of *OsUSP*s ranged from 222 bp (*OsUSP31*) to 2817 bp (*OsUSP20*) long. Altogether, the forty-four *OsUSP* genes encoded for a total of 57 proteins. The protein length ranged from 74 aa (OsUSP31) to 939 aa (OsUSP20). Meanwhile, the protein weight varied greatly from the lowest of 17.22 kDa (OsUSP13) to the highest of 120.81 kDa (OsUSP41), and the predicted isoelectric point range from 4.74 (OsUSP30) to 10.39 (OsUSP37).

In silico subcellular localization prediction indicated that *OsUSPs* were localized in the cytoplasm, mitochondria, nucleus, and chloroplast (**Table 1**).

Analysis of *OsUSP* Gene Structure, Conserved Motifs and Amino Acid Content

To better understand the development of the *OsUSP* gene family, the exon–intron distribution of *OsUSP* genes was analyzed. In general, the number of exons and introns in *OsUSP* genes showed moderate variation (**Figure 2A**). In the case of introns, the numbers ranged from zero to ten, with *OsUSP32* having the highest and *OsUSP23* having the lowest number of introns (10 and 0, respectively). On average, *OsUSP* genes that contained both USP and kinase domains had a higher number of introns compared to those with only a USP domain. This indicates a similarity in the exon–intron architecture between genes that were more similar on a protein level. Similarly, the number of exons varied from one to eleven, with *OsUSP32* having the highest number of exons. Only a single exon was found in four genes, namely – *OsUSP19, OsUSP23, OsUSP26,* and *OsUSP30*. Several *OsUSP* genes such as *OsUSP1, OsUSP12, OsUSP25, OsUSP31* contain multiple 3′-UTR regions. This likely reflects the possibility of alternative splicing in these genes.

Elucidation of conserved motifs using MEME software revealed a total of 10 conserved motifs among rice USP proteins (**Figure 2B**). As expected, there was quite some variation in terms of motif number in each protein. In general, most proteins either contained three, nine or all 10 of the conserved motifs. This observation perfectly coincided with the presence of all three domains – USP, Kinase, and U-Box; two domains – USP and Kinase; or just a single USP domain in the protein. In the former case, 10 motifs were found in the proteins, whereas for those with only the USP domain, the number of conserved motifs was between one to three. From this observation, it could be deduced that motifs 3, 4, and 7 are associated with the USP domain; motifs 1, 2, 5, 6, 8, and 9 are linked to the kinase domain and motif 10 corresponds to the U-box domain. In addition to this, the amino acid (aa) composition of OsUSPs (**Figure 2C**) showed that OsUSPs with the same domain(s) show more similarity in their aa pattern compared to those with slightly different domain attributes. This can be exemplified by looking at the aa composition of USP 32-39 (**Figure 2C**). Here, USP 32-35 and 38-39 have both a USP and a kinase domain and are relatively alike in their aa composition. However, USP36 and USP35 lack a kinase domain and it is discernible that their aa composition is more diverse from USP 32-35 and 38-39. There appears to be a consistency in gene structure, motif conservation, and AA composition with the type of domain a given OsUSP contains.

Chromosomal Distribution and Synteny Analysis

The identified *OsUSP* genes were distributed over 11 out of 12 rice chromosomes (**Figure 3A**). The exact number of *OsUSP* genes per chromosome varied widely. Chromosome 2 contained the highest number of eight *USP* genes, followed by chromosome 1 with seven genes. A relatively moderate number of *USP* genes were found in chromosomes 5 and 12, where the number reached six and five, respectively. Both chromosomes 3 and 10 contained four members. As for other chromosomes, the number of *USP* gene family members ranged from one to three, with chromosome 4 being the only exception, which did not contain any *USP* genes.

The gene duplication events among *OsUSP* genes were investigated to evaluate the extension of this gene family members. A total of 11 *OsUSP* gene pairs were found to be duplicated- *OsUSP1* and *OsUSP22, OsUSP2* and *OsUSP23, OsUSP2* and *OsUSP17, OsUSP4* and *OsUSP29, OsUSP5* and *OsUSP26, OsUSP6* and *OsUSP25, OsUSP7* and *OsUSP24, OsUSP9* and *OsUSP29, OsUSP15* and *OsUSP28, OsUSP18* and *OsUSP31*, and *OsUSP2* and *OsUSP23* (**Supplementary Table 6**). All of them are possessing segmental type of duplication, except one with tandem type. A negative selection was observed for all the duplicated pairs except one with the ratio of nonsynonymous (Ka) to synonymous (Ks) values more than one. The calculated divergence time of the duplication was varied from 28 to 150 Mya (**Supplementary Table 6**).

Chromosomal synteny analysis was carried out to understand the expansion and diversification of *USP* genes within the rice genome and to compare with other monocot (maize) and dicot (*Arabidopsis*) organisms (**Figure 3B**). Both segmental and tandem duplication were visible for *USP* genes within the rice genome, with the former dominating the latter. A total of sixteen paralogous *USP* genes were recognized within the rice genome that were localized on chromosomes 1, 2, 5, 6, 11, and 12. Whereas, a single tandem duplication was detected on chromosome 12 between *OsUSP43* and *OsUSP44*. Interesting observations came to light during the syntenic comparison of *OsUSP* genes with *Arabidopsis* and maize. Starting with the dicot genome, only four *OsUSP* orthologs were detected in the *Arabidopsis* genome and were limited to chromosomes AtChr3 and AtChr5 (**Figure 3B**). In contrast to this, a relatively large number of *OsUSP* orthologs were detected in the monocot plant, maize. In total, 37 orthologous genes were spread throughout all maize chromosomes (**Figure 3B**). This likely indicates that *OsUSP* genes have diverged more from their counterparts in the *Arabidopsis* genome compared to the more closely related monocot, maize. Overall, this indicates that both duplication and intense diversification have played a characteristic role in the evolution of the *OsUSP* gene family.

Phylogeny of USPs From Three Species and Their Domain Architecture

To explore the evolutionary pattern of USP proteins in three species – rice, *Arabidopsis*, and *E. coli*, we carried out a phylogenetic analysis using their all USP members (**Figure 4**). The addition of domain architecture of these proteins to the tree showed two clear clusters – one with members that contain only a USP domain (**Figure 4**, green cluster) and the other with members that mostly contain an additional kinase domain on top of the USP domain (**Figure 4**, yellow cluster). Overall, the number of proteins with only a USP domain was higher than the

TABLE 1 | List of identified *OsUSP* genes along with their molecular attributes.
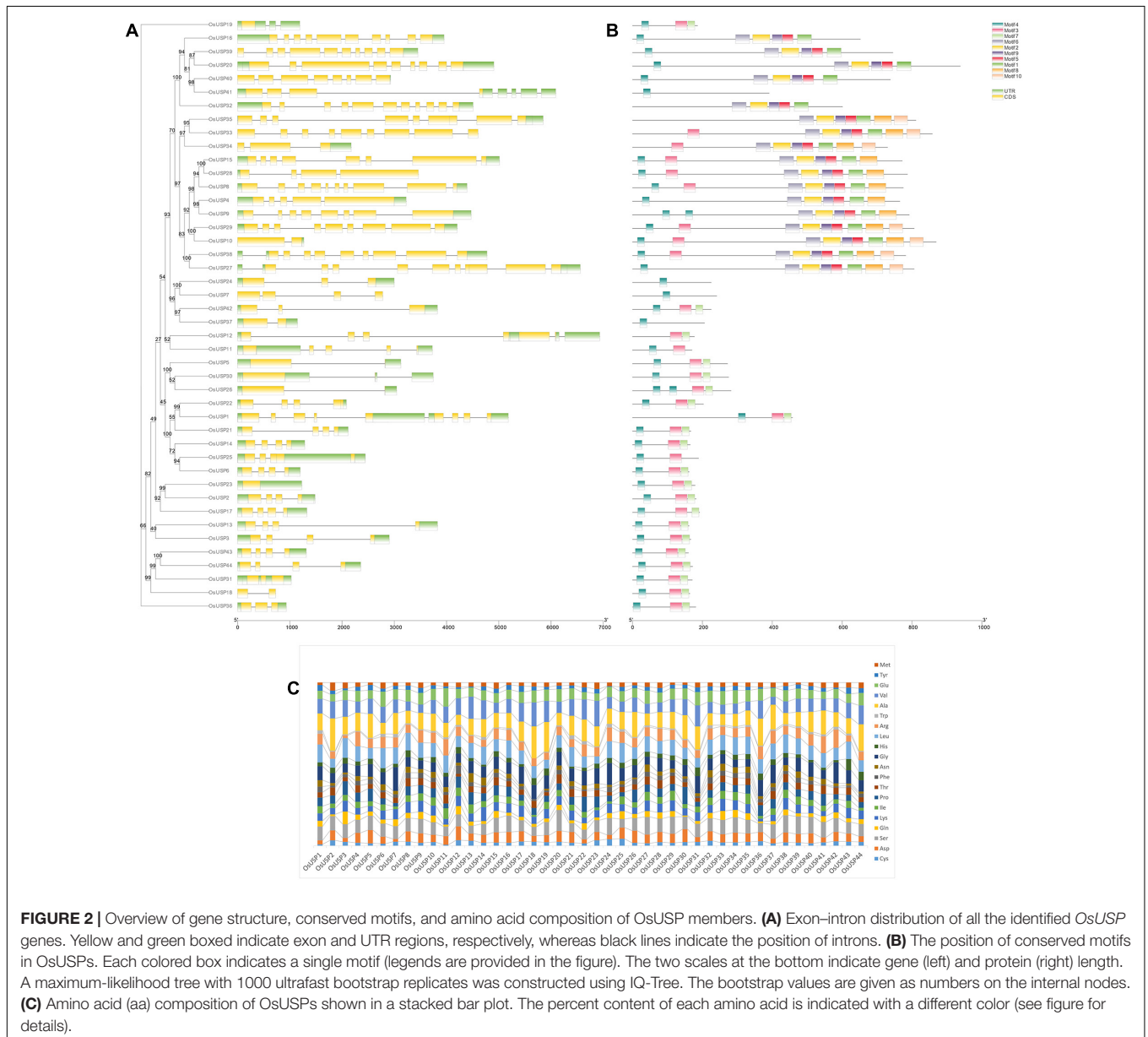
Gene	Chr. no	Locus	Transcripts	CDS coordinate (5' to 3')	CDS (bp)	PP (aa)	Mass (kDa)	pI	Localization
OsUSP1	Chr1	Os01g0170600	Os01t0170600-01	3645508–3648396	819	273	17.95	10.13	Mt[1,2], Chl[1,2,3]
			Os01t0170600-02	3645512–3650691	483	161	5.01	10.13	
OsUSP2	Chr1	Os01g0303800	Os01t0303800-01	11245253–11246737	549	183	18.76	6.22	Cyt[1,2], Chl[1], Nu[1]
			Os01t0303800-02	1245276–11246714	549	183	18.76	6.22	
OsUSP3	Chr1	Os01g0511100	Os01t0511100-01	17990846–17993750	504	168	17.92	7.55	Chl[1], Cyt[1,2], Mt[1]
OsUSP4	Chr1	Os01g0581400	Os01t0581400-01	22538998–22542226	2298	766	84.03	7.98	Cyt[1,2], Chl[1], Nu[1]
OsUSP5	Chr1	Os01g0783500	Os01t0783500-01	33198608–33201733	789	263	28.79	4.99	Chl[1,3], Mt[1]
OsUSP6	Chr1	Os01g0849600	Os01t0849600-01	36496934–36498132	489	163	18.03	6.52	Cyt[1,2], Mt[1,2], Chl[1,2]
OsUSP7	Chr1	Os01g0875300	Os01t0875300-00	37982651–37985429	945	315	25.28	9.99	Chl[1,2,3], Cyt[2], Nu[2]
OsUSP8	Chr2	Os02g0152300	Os02t0152300-01	2869670–2874062	2328	776	86.19	6.66	Nu[1,2], Chl[1], Cyt[1,2], Mt[1,2]
OsUSP9	Chr2	Os02g0218400	Os02t0218400-01	6609459–6613930	2193	731	88.49	8.75	Nu[1,2], Cyt[1,2], Mt[2]
OsUSP10	Chr2	Os02g0218600	Os02t0218600-00	6621779–6623047	1092	364	95.56	8.47	Chl[1,3], Mt[2], Nu[2], Cyt[2]
OsUSP11	Chr2	Os02g0705400	Os02t0705400-01	29113442–29114646	258	86	19	5.37	Cyt[1], Mt[2], Chl[2]
			Os02t0705400-02	29113445–29117170	555	185	19	5.37	
OsUSP12	Chr2	Os02g0707900	Os02t0707900-01	29260063–29265497	534	178	19.87	6.39	Chl[1,2], Cyt[1,2], Nu[1,2], Mt[1,2]
			Os02t0707900-02	29260085–29265446	534	178	19.87	6.39	
			Os02t0707900-03	29258565–29260280	597	199	19.87	6.39	
OsUSP13	Chr2	Os02g0760500	Os02t0760500-01	32023322–32027150	489	163	17.22	6.75	Cyt[1,2], Mt[2], Chl[2]
OsUSP14	Chr2	Os02g0773200	Os02t0773200-01	32643551–32644836	498	166	17.88	8.82	Cyt[1], Mt[2], Chl[1]
OsUSP15	Chr2	Os02g0787200	Os02t0787200-01	33434089–33439102	2319	773	84.75	7.37	Nu[1,2], Chl[1], Cyt[1,2]
OsUSP16	Chr3	Os03g0241600	Os03t0241600-01	7483344–7487296	1959	653	70.89	7.41	Chl[1,2], Nu[1], Cyt[1,2], Mt[1]
			Os03t0241600-02	7483349–7484709	762	254	43.47	9.92	
OsUSP17	Chr3	Os03g0305400	Os03t0305400-01	10839459–10840785	543	181	18.93	4.87	Chl[1], Mt[2], Cyt[1,2]
OsUSP18	Chr3	Os03g0344166	Os03t0344166-00	12820005–12820732	321	107	16.84	6.62	Cyt[1,2], Mt[1,2], Chl[2]
OsUSP19	Chr3	Os03g0750000	Os03t0750000-01	30901840–30903031	258	86	19.23	6.95	Cyt[1,2], Chl[1,2], Mt[2]
OsUSP20	Chr3	Os03g0839900	Os03t0839900-01	35312754–35317359	2817	939	101.27	4.99	Nu[1,2], Mt[1]
			Os03t0839900-02	35312453–35317266	2817	939	101.27	4.99	
OsUSP21	Chr5	Os05g0157200	Os05t0157200-01	3345130–3347246	504	168	18	5	Cyt[1], Chl[1]
OsUSP22	Chr5	Os05g0170200	Os05t0170200-01	4205610–4207693	390	130	21.65	10.08	Chl[1,2], Cyt[1,2], Mt[1,2]
			Os05t0170200-02	4205610–4207693	612	204	21.65	10.08	
OsUSP23	Chr5	Os05g0355400	Os05t0355400-01	16859727–16860956	333	111	18.71	7.83	Mt[2], Chl[1,2], Cyt[1,2]
OsUSP24	Chr5	Os05g0428400	Os05t0428400-00	21006679–21009677	678	226	24.08	10.01	Nu[1,2], Chl[1], Mt[2]
OsUSP25	Chr5	Os05g0453700	Os05t0453700-01	22276282–22278726	570	190	20.9	6.23	Cyt[1], Chl[1], Nu[1], Mt[2]
			Os05t0453700-02	22276284–22278726	498	166	18.53	6.51	
OsUSP26	Chr5	Os05g0501700	Os05t0501700-01	24696226–24699271	807	269	29.82	5.74	Chl[1,2,3]
OsUSP27	Chr6	Os06g0140800	Os06t0140800-01	2140055–2146616	2421	807	90.05	6.58	Nu[1,2], Cyt[1], Chl[1]
			Os06t0140800-02	2140306–2142183	1281	427	90.05	6.58	
OsUSP28	Chr6	Os06g0191900	Os06t0191900-00	4641462–4644922	2436	812	86.41	8.01	Nu[1,2], Chl[1], Cyt[1,2], Mt[2]
OsUSP29	Chr6	Os06g0574200	Os06t0574200-01	22271338–22275542	2421	807	88.28	8.11	Chl[1,2,3], Nu[1,2], Mt[1,2]
OsUSP30	Chr7	Os07g0551400	Os07t0551400-01	21894865–21898610	807	269	28.9	4.83	Chl[1,2,3], Cyt[1], Nu[1,2], Mt[1]
			Os07t0551400-02	21897237–21898566	807	269	28.19	4.74	
OsUSP31	Chr7	Os07g0673400	Os07t0673400-01	28468527–28469554	516	172	17.96	7.01	Cyt[1,2], Chl[1,2], Nu[1], Mt[2]
			Os07t0673400-02	28468619–28469554	222	74	17.96	7.01	
OsUSP32	Chr8	Os08g0249100	Os08t0249100-01	9101238–9105746	1806	602	67.33	6.51	Cyt[1,2], Chl[1,2], Nu[1,2]
OsUSP33	Chr9	Os09g0569800	Os09t0569800-00	22736559–22741162	2577	859	96.14	6.79	Chl[1,3], Nu[2], Cyt[2]
OsUSP34	Chr9	Os09g0570000	Os09t0570000-01	22749994–22752169	1083	361	81.92	5.98	Cyt[1,2], Nu[1,2]
OsUSP35	Chr10	Os10g0100500	Os10t0100500-01	67119–72969	2436	812	90.78	7.15	Chl[1,3], Nu[1,2], Cyt[1,2], Mt[2]
			Os10t0100500-02	71174–72765	972	324	90.78	7.15	
OsUSP36	Chr10	Os10g0437500	Os10t0437500-01	15663637–15664570	546	182	18.88	5.87	Cyt[1,2], Chl[1,2]
OsUSP37	Chr10	Os10g0463300	Os10t0463300-01	17074150–17075296	621	207	21.48	10.39	Chl[1,2], Cyt[1], Nu[1,2]
OsUSP38	Chr10	Os10g0561500	Os10t0561500-01	22136957–22141731	2349	783	86.88	7.06	Cyt[1,2], Chl[1,3], Nu[1,2]
OsUSP39	Chr11	Os11g0194900	Os11t0194900-01	4734980–4738431	2004	668	80.54	7.66	Chl[1,2,3], Cyt[1,2]
OsUSP40	Chr12	Os12g0180500	Os12t0180500-00	4056054–4058982	2190	730	78.57	7.48	Chl[1,2], Mt[1,2], Cyt[1]

(Continued)

Gene	Chr. no	Locus	Transcripts	CDS coordinate (5′ to 3′)	CDS (bp)	PP (aa)	Mass (kDa)	pI	Localization
OsUSP41	Chr12	Os12g0181200	Os12t0181200-00	4091707–4097797	1176	392	120.81	7.97	Chl[1], Mt[1,2], Cyt[1], Nu[2]
OsUSP42	Chr12	Os12g0501400	Os12t0501400-01	19089135–19092961	678	226	24.24	10.28	Chl[1,2], Nu[2]
OsUSP43	Chr12	Os12g0552400	Os12t0552400-01	22427863–22429178	483	161	17.54	7.36	Cyt[1,2], Chl[1], Nu[2], Mt[2]
OsUSP44	Chr12	Os12g0552500	Os12t0552500-01	22429432–22431789	510	170	18.42	7.28	Cyt[1,2], Chl[1,2], Nu[1,2], Mt[2]

[1]WoLFPSORT, [2]CELLO subcellular localization prediction tool, [3]ChloroP.
Mt, mitochondria; chl, chloroplast; Nu, nucleus; Cyt, cytoplasm.

FIGURE 2 | Overview of gene structure, conserved motifs, and amino acid composition of OsUSP members. **(A)** Exon–intron distribution of all the identified *OsUSP* genes. Yellow and green boxed indicate exon and UTR regions, respectively, whereas black lines indicate the position of introns. **(B)** The position of conserved motifs in OsUSPs. Each colored box indicates a single motif (legends are provided in the figure). The two scales at the bottom indicate gene (left) and protein (right) length. A maximum-likelihood tree with 1000 ultrafast bootstrap replicates was constructed using IQ-Tree. The bootstrap values are given as numbers on the internal nodes. **(C)** Amino acid (aa) composition of OsUSPs shown in a stacked bar plot. The percent content of each amino acid is indicated with a different color (see figure for details).

other type with more structural diversity in terms of their domain architecture. Focusing on the yellow cluster, most members had a kinase domain in addition to the characteristic USP domain. For some members, the presence of a U-box domain can also be seen. Interestingly, two rice USPs that lack a kinase domain (OsUSP10

and 41) were also found to group in this cluster. With respect to comparison of USP evolution between species, all *E. coli* USPs grouped together in the first cluster (green). A closer inspection of the tree showed that USPs usually clustered closest to a paralog from the same species rather than its ortholog from another
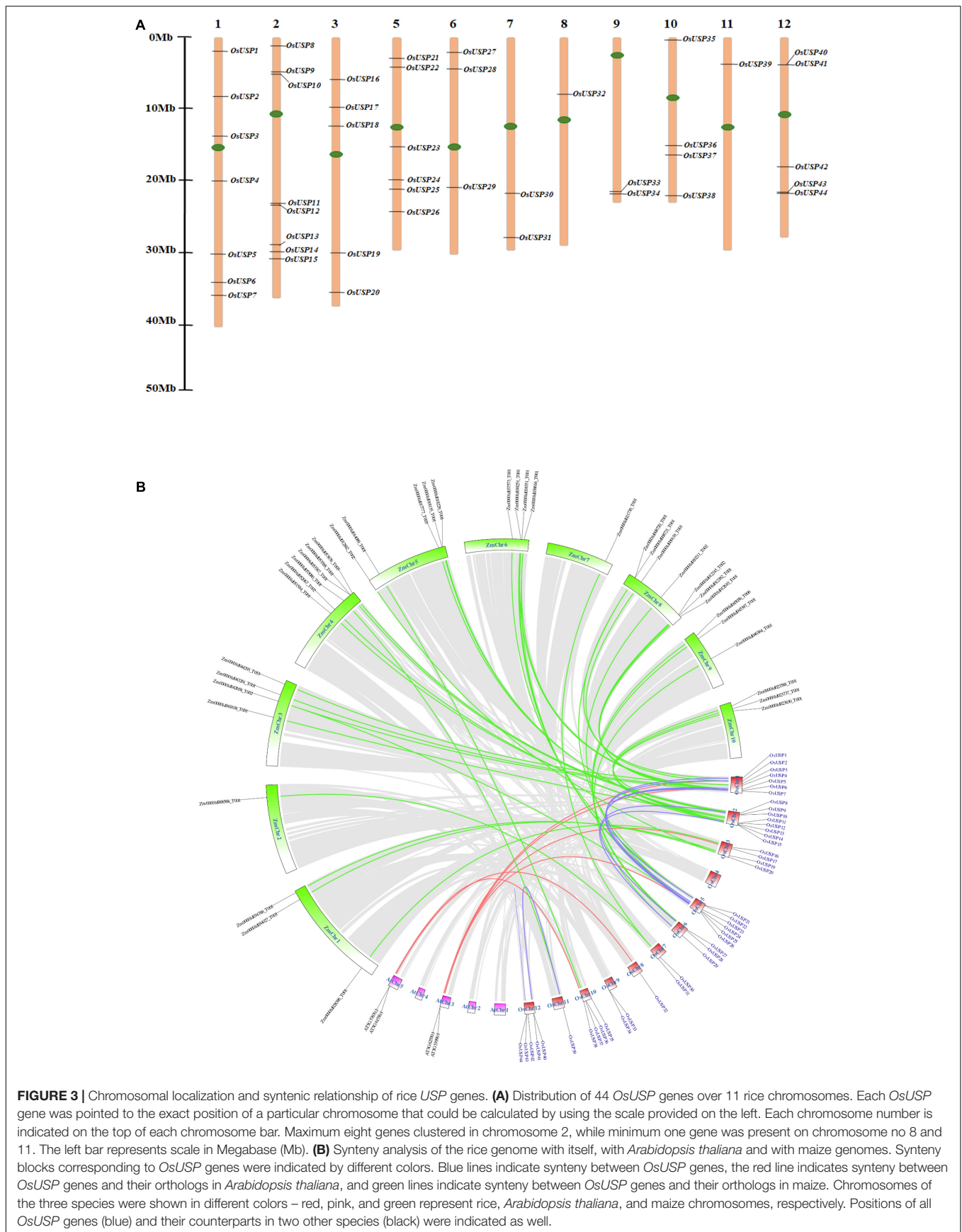
FIGURE 3 | Chromosomal localization and syntenic relationship of rice *USP* genes. **(A)** Distribution of 44 *OsUSP* genes over 11 rice chromosomes. Each *OsUSP* gene was pointed to the exact position of a particular chromosome that could be calculated by using the scale provided on the left. Each chromosome number is indicated on the top of each chromosome bar. Maximum eight genes clustered in chromosome 2, while minimum one gene was present on chromosome no 8 and 11. The left bar represents scale in Megabase (Mb). **(B)** Synteny analysis of the rice genome with itself, with *Arabidopsis thaliana* and with maize genomes. Synteny blocks corresponding to *OsUSP* genes were indicated by different colors. Blue lines indicate synteny between *OsUSP* genes, the red line indicates synteny between *OsUSP* genes and their orthologs in *Arabidopsis thaliana*, and green lines indicate synteny between *OsUSP* genes and their orthologs in maize. Chromosomes of the three species were shown in different colors – red, pink, and green represent rice, *Arabidopsis thaliana*, and maize chromosomes, respectively. Positions of all *OsUSP* genes (blue) and their counterparts in two other species (black) were indicated as well.

FIGURE 4 | Phylogenetic relationship between rice, *Arabidopsis thaliana*, and *E. coli* USPs. A maximum-likelihood tree with 1000 ultrafast bootstrap replicates was produced using IQ-Tree. The final tree was visualized and edited in iTOL. USP members from the three different species are specified using different colored shapes – blue rectangles, red circles, and green stars sequentially specify *Arabidopsis thaliana*, *Oryza sativa*, and *E. coli* USPs. Overall, USPs were sub-grouped into two clusters – green and yellow based on the domain architecture of each protein. The green clusters exclusively contain protein that harbors only a USP domain. As for the yellow cluster, almost all members harbor an additional kinase or U-Box domain or both (except a few). Information regarding the presence and position of a domain was added in iTOL and the name of each domain is indicated within the corresponding boxes.

species. This is because only in a relatively small number of cases can clear orthologs between the two species be identified in the tree in **Figure 4**. For instance, clear orthologs in *Arabidopsis* could only be discerned for rice USPs – 9, 11, 14, 17, 25, and 37.

In silico Identification of SSRs, and Prediction of Glycosylation and Phosphorylation Sites

Molecular markers allow the identification of genes of interest within a particular location. A major concern of genome

analysis is to explore molecular markers related to genetic factors underlying observable traits. Forty-four *OsUSP* genes with all splice variants were analyzed for SSR markers. Out of 61 sequences, 19 SSR markers were distributed among 15 sequences. The most abundant were trinucleotide repeats accounting for 89.47% (17 occurrences), followed by mononucleotide and dinucleotide repeats both with one occurrence accounting for 5.26% of all SSRs, no tetra, penta or hexanucleotide repeats were found. The dominant SSR motif was GCG with a frequency of 47.36 with nine occurrences followed by other mono, di and trinucleotide repeats (**Supplementary Table 7**). Higher

repeat numbers were observed in trinucleotide SSRs. Several sequences possessing more than one SSR marker were found three. Furthermore, glycosylation plays a key role in secondary protein processing for the functioning of protein within cells and in determining protein structure, function, and stability. In this study, 25 out of 44 OsUSPs have shown one to many predicted glycosylation sites with the threshold of 0.5. A total of 68 glycosylation sites have been predicted in 25 OsUSP proteins. The highest number of glycosylation sites was found in OsUSP8 with nine sites followed by OsUSP10, OsUSP32, OsUSP35 (five sites) and OsUSP27, OsUSP33, OsUSP38 with four sites. *N*-Glycosylation score greater than 0.5 and jury agreement 9/9 or potential > 0.75 indicates high specificity of glycosylation site. Among the total of 68 predicted sites, 19 sites in 13 OsUSP proteins have the potential score greater than 0.5 and jury agreement 9/9, and thus, have the highest chance of glycosylation mediated stable structure formation (**Supplementary Table 8**). Protein phosphorylation is the another most significant type of post-translational modification of cells in which protein kinases phosphorylate an amino acid residue, mostly Tyr, Thr, Ser in the case of eukaryotes by adding covalently bound phosphate groups. A significant number of phosphorylation sites were predicted in all the members of OsUSP (**Supplementary Table 9**). The highest number of phosphorylation sites were predicted in OsUSP20, followed by OsUSP29, OsUSP27 and OsUSP39. Predicted results also showed that serine (S) sites were more phosphorylated than threonine (T) and tyrosine (Y) sites. As for the types of kinases, the prevalence of the predicted ones was not specific, while PKC was the most common.

Analysis of Expression Pattern of *OsUSP* Genes During Developmental Stages and in Various Tissues of Rice

To get a notion about the magnitude of *OsUSP* genes expression, we carried out an expression analysis using previously generated transcriptomics data for *OsUSP* genes during different developmental stages (**Figure 5A**) and anatomical tissues (**Figure 5B**). Most of the *OsUSP* genes consistently show high levels of expression during all the developmental stages (**Figure 5A**) except *OsUSP18, OsUSP40, OsUSP41,* and *OsUSP42.* These genes show relatively low expression levels in almost all stages. However, some interesting candidates show preferentially high expression during certain stages. For instance, *OsUSP15* is highly expressed during the flowering and heading stage but less prominent during germination, seedling, tillering, and mature grain stage. A similar observation is true for *OsUSP8* and *OsUSP37* with the latter showing a stronger contrast in expression.
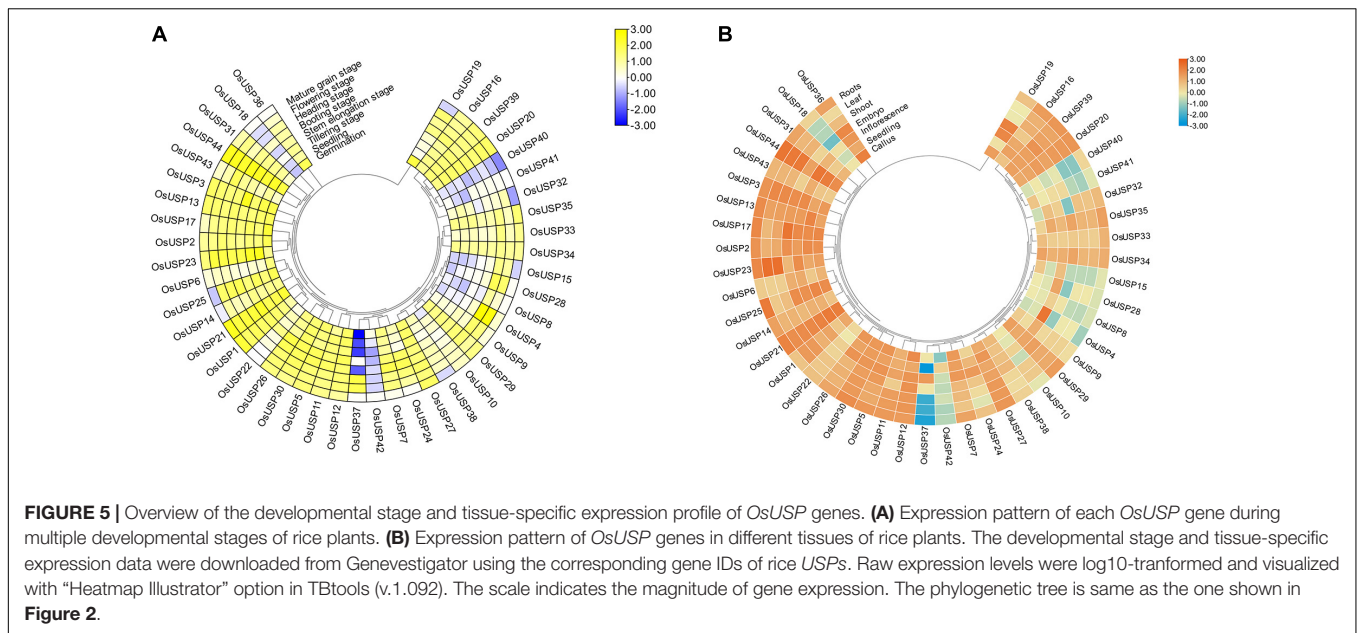
Moving onto expression profiles in tissues, a similar pattern was observed where most *OsUSP* genes exhibited moderate to high expression levels in all tissues (**Figure 5B**), with exception of *OsUSP4, OsUSP8, OsUSP15, OsUSP18, OsUSP28, OsUSP37, OsUSP40, OsUSP41,* and *OsUSP42.* Some genes showed a mixed pattern, while some tissues are showing higher expression for few genes than others. *OsUSP4* showed very high expression in inflorescence tissues but moderate to low level

of expression in others. Comparably, *OsUSP37* had a very low expression in roots, shoots, leaves, and seedlings but a high expression in inflorescence tissues. Overall, most of *OsUSP* genes showed an appreciable level of expression both during different developmental stages and tissues of the plant.

Analysis of Stress-Specific Expression Pattern of *OsUSP* Genes and Their *Cis*-Regulatory Elements

To evaluate the role of *OsUSP*s in stress conditions, analysis of the expression data of all the identified *OsUSP* genes were performed under multiple stress conditions including – heat, salt, drought, dehydration, and submergence (**Figure 6A**). Expectedly, most *OsUSP* genes showed strong upregulation or downregulation in response to these adversities. However, the number of upregulated genes was slightly higher than those that were downregulated. Some *USP* genes including – *OsUSP2, OsUSP17, OsUSP22, OsUSP23, OsUSP26, OsUSP31,* and *OsUSP36* were strongly upregulated (even higher than 6-fold) under submergence conditions. All genes (except *OsUSP26*) showed a strong upregulation from the early onset of submergence. Some genes such as *OsUSP1, OsUSP10, OsUSP16, OsUSP20, OsUSP28, OsUSP42* showed downregulation as well. However, the magnitude of downregulation was not as strong for those that were upregulated. In the case of heat stress, the expression of *OsUSP* genes varied quite a lot. The strongest upregulation was detected for *OsUSP14*, whereas some genes showed both upregulation and downregulation in a temporally separated fashion like *OsUSP15, OsUSP22, OsUSP30*, etc. Moving on to dehydration and drought stress, upregulation of several USP genes was more prominent at the early onset of stress application. For instance, genes *OsUSP3, OsUSP5, OsUSP11, OsUSP14, OsUSP19, OsUSP22, OsUSP23, OsUSP26,* and *OsUSP36* showed considerable upregulation at the earlier stages of drought stress, but their expression varied in the later time points. Some of these genes such as *OsUSP5* and *OsUSP14*, showed upregulation in both conditions and at all time points. There was also a mixed response of *OsUSP* genes in response to salt, with some being upregulated and others downregulated. The magnitude of upregulation was stronger mostly at the longer period of salt application (24 h). In terms of consistency across all stress conditions, the expression profile of *OsUSP2, OsUSP3, OsUSP22, OsUSP32, OsUSP33, OsUSP39,* and *OsUSP44* was very interesting. In general, *OsUSP3, OsUSP22, OsUSP39,* and *OsUSP44* were mostly upregulated and *OsUSP2, OsUSP32,* and *OsUSP33* were mostly downregulated across the conditions (**Figure 6A**).

Additionally, we checked the composition of the putative promoter region of *OsUSP* genes concerning the presence of different *cis*-regulatory elements. After surveying the 1000 bp-upstream sequence of each gene, most rice *USP* promoters harbored numerous *cis*-regulatory elements which are known to be involved in stress-induced gene regulation (**Figure 6B**). The motifs that were most commonly found in the *OsUSP* promoter regions were – ABRE, ARE, LTR, MBS, and TC-rich repeats, which are recognized for attracting stress-specific transcription

FIGURE 5 | Overview of the developmental stage and tissue-specific expression profile of *OsUSP* genes. **(A)** Expression pattern of each *OsUSP* gene during multiple developmental stages of rice plants. **(B)** Expression pattern of *OsUSP* genes in different tissues of rice plants. The developmental stage and tissue-specific expression data were downloaded from Genevestigator using the corresponding gene IDs of rice *USPs*. Raw expression levels were log10-tranformed and visualized with "Heatmap Illustrator" option in TBtools (v.1.092). The scale indicates the magnitude of gene expression. The phylogenetic tree is same as the one shown in **Figure 2**.

factors (TFs). To add to this line of evidence, a GO enrichment analysis on all the *OsUSP* genes was performed (**Figure 6C**). This showed that out of the 44 genes identified, 41 *OsUSP* genes were annotated to be involved in stress response (GO:0006950). In short, the stress-specific modulations of rice *USP* genes are clear.

Validation of Gene Expression Profile of Nine Selected *OsUSP* Genes in Response to Abiotic Stress
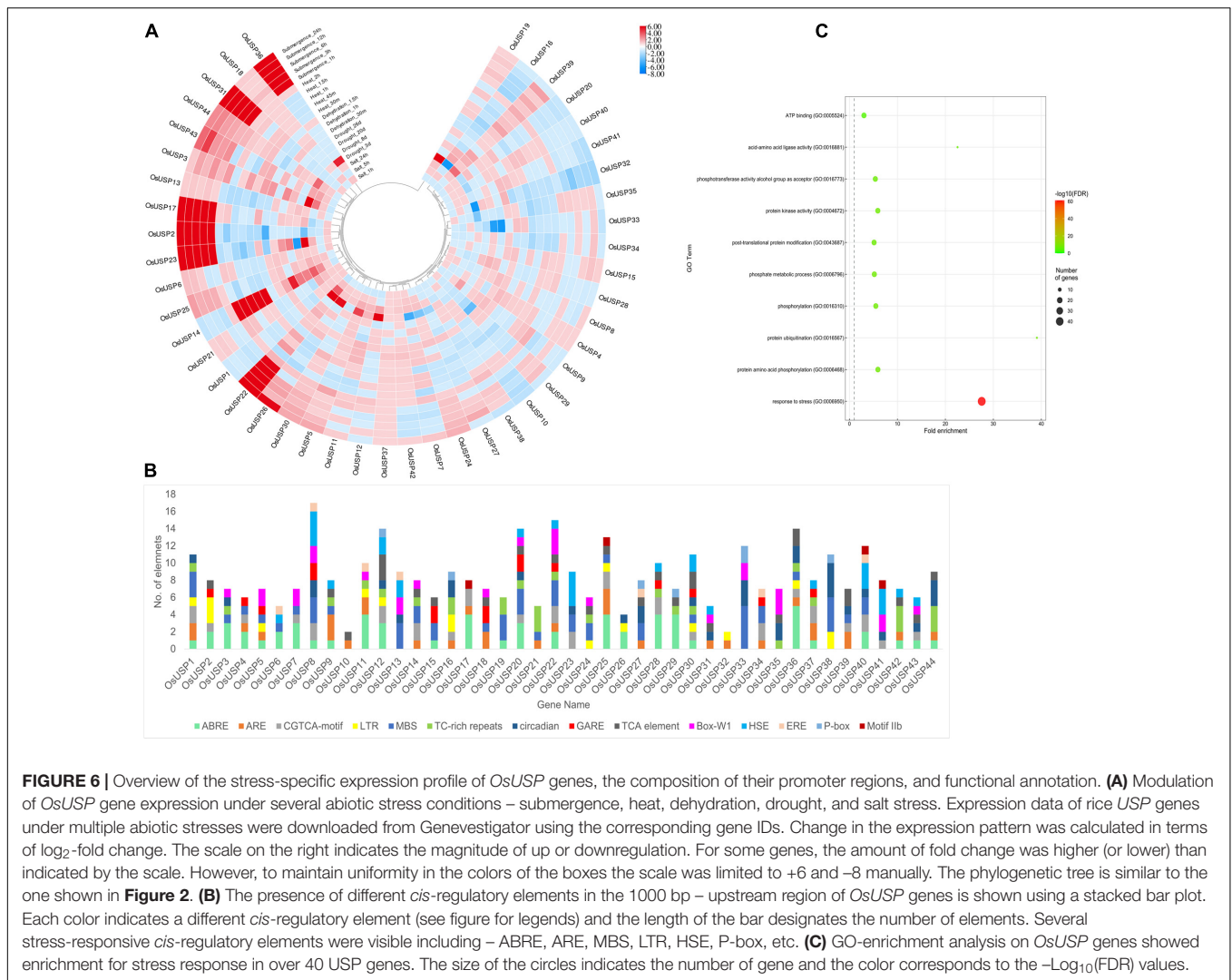
As an expression profile of genes under stress condition is important to understand their function, it is crucial to assess mRNA-seq expression patterns experimentally to ensure consistency. Thus, we carried out quantitative RT-PCR analysis of nine selected *OsUSP* genes in leaf tissue after 16 h of abiotic stress treatments including salt, drought, cold, heat, dehydration, H_2O_2 and hormonal treatment – ABA, SA, and GA3. The real-time PCR expression profile of the selected genes (**Figure 7**) reveal a good correlation with the mRNA-seq data (**Figure 6A**). Among nine genes, *OsUSP2, OsUSP3, OsUSP6, OsUSP12, OsUSP14, OsUSP32, OsUSP33* showed considerable upregulation in heat stress whereas *OsUSP22* and *OsUSP36* showed downregulation. Consistent with the mRNA-seq expression data, *OsUSP14* and *OsUSP22* genes showed a similar pattern of expression under heat stress. Under salinity stress *OsUSP2, OsUSP3, OsUSP6, OsUSP12, OsUSP14* and *OsUSP32* showed low to medium upregulation while *OsUSP22, OsUSP33* and *OsUSP36* showed downregulation in rice leaves. Some genes such as, *OsUSP2, OsUSP22* and *OsUSP32* showed differential expression patterns in salt stress which is different from the mRNA-seq data. However, the expression level of *OsUSP3* and *OsUSP33* were consistent with the expression pattern mRNA-seq data. Furthermore, the qRT-PCR results revealed that the expression level of *OsUSP2, OsUSP3, OsUSP12, OsUSP14, OsUSP32, OsUSP33* was upregulated under drought stress. Apart from that, *OsUSP6*

showed no fold change, compared to control under drought stress. In dehydrated condition, only *OsUSP2, OsUSP12* and *OsUSP33* showed reasonable upregulation, while *OsUSP14* showed downregulation. The occasional inconsistency of gene expression of the selected genes between mRNA-seq and RT-PCR might be due to the differences in sampling time following stress/hormone treatment or the genotypic difference between the rice japonica variety and the BRRI-53 variety.

Expression analysis of nine selected *OsUSP* genes were further conducted in other abiotic treatments such as cold, oxidative, and hormonal stress. In cold treatment, *OsUSP* genes showed low to medium upregulation in leaf samples. *OsSUP6* does not show any expression in drought stress but is highly expressed in cold stress whereas *OsUSP33* showed negligible up-regulation which is close to control. *OsUSP14, OsUSP32, OsUSP36* showed up-regulation under cold stress. But we observed that the other *USP* genes such as *OsUSP2, OsUSP3, OsUSP12, OsUSP22* and *OsUSP36* were downregulated in cold. *OsUSPs* showed down-regulation in oxidative stress except *OsUSP3* and *OsUSP12*, which need to be investigated further. Furthermore, *OsUSP* showed a differential regulation in rice leaf tissue under hormonal influences like ABA, SA and GA3 (**Figure 7**). The qRT-PCR results suggested that most of the *USP* genes such as *OsUSP2, OsUSP3, OsUSP6, OSUSP14* and *OsUSP32* up-regulated in all the hormonal stresses. Apart from that, *OsUSP12* and *OsUSP33* showed upregulation but downregulated in ABA stress. The varying responses of *OsUSPs* could thus be attributed to different abiotic and hormonal stresses.

Molecular Docking Analysis of Selected OsUSPs

Among 44 OsUSP proteins, 9 OsUSPs showed significant up and down-regulation to various stresses and were selected as receptors for docking study to check their affinity toward various
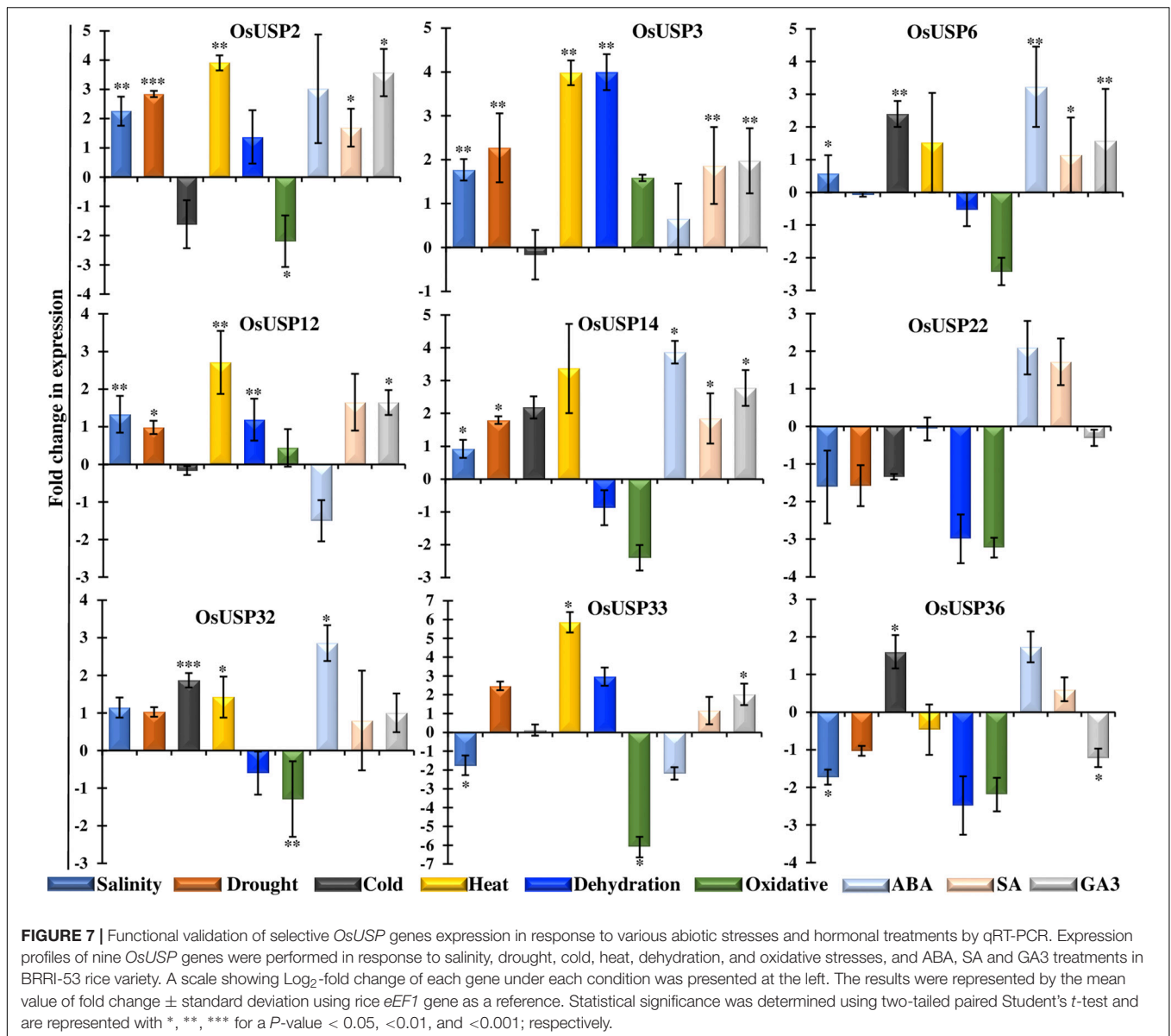
FIGURE 6 | Overview of the stress-specific expression profile of *OsUSP* genes, the composition of their promoter regions, and functional annotation. **(A)** Modulation of *OsUSP* gene expression under several abiotic stress conditions – submergence, heat, dehydration, drought, and salt stress. Expression data of rice *USP* genes under multiple abiotic stresses were downloaded from Genevestigator using the corresponding gene IDs. Change in the expression pattern was calculated in terms of $\log_2$-fold change. The scale on the right indicates the magnitude of up or downregulation. For some genes, the amount of fold change was higher (or lower) than indicated by the scale. However, to maintain uniformity in the colors of the boxes the scale was limited to +6 and –8 manually. The phylogenetic tree is similar to the one shown in **Figure 2**. **(B)** The presence of different *cis*-regulatory elements in the 1000 bp – upstream region of *OsUSP* genes is shown using a stacked bar plot. Each color indicates a different *cis*-regulatory element (see figure for legends) and the length of the bar designates the number of elements. Several stress-responsive *cis*-regulatory elements were visible including – ABRE, ARE, MBS, LTR, HSE, P-box, etc. **(C)** GO-enrichment analysis on *OsUSP* genes showed enrichment for stress response in over 40 USP genes. The size of the circles indicates the number of gene and the color corresponds to the –$\log_{10}$(FDR) values.

inhibitors (**Table 2**). Of these 9, only two (OsUSP32, OsUSP33) have both USP and kinase domain, docked against UspA and PKC inhibitors. Rest seven contain only the USP domain, therefore, they were docked against only UspA inhibitors. All these selected nine proteins showed significant affinity toward UspA inhibitors (ZINC000104153710, ZINC000004268284, ZINC000000217308), and both OsUSP32 and OsUSP33 showed affinity toward four kinase inhibitors (Fisetin, Luteolin, Myricetin, Quercetin) (**Table 2**). Among the three USP inhibitors the highest affinity was found for Zinc000104153710 and in the case of four kinase inhibitors, Luteolin showed the highest affinity. Among all nine selected proteins, OsUSP32 showed the highest affinity toward UspA inhibitor Zinc000104153710 and PKC inhibitor luteolin with a binding energy of −9.6 kcal/mol and −9.2 kcal/mol, respectively, followed by OsUSP33 (−8.5 kcal/mol, −8.2 kcal/mol, respectively). The protein-ligand interaction of docked complexes was visualized using the Discovery Studio program (**Figures 8A–D**). The best-scored protein OsUSP32 with the highest binding affinity toward Zinc000104153710 formed one hydrogen bond with Asp443

having a distance of 2.21 Å, two pi-sigma bond with Val324 (3.93 Å), Leu432 (3.88 Å), one pi-alkyl bond with Ala359 (5.44 Å) and five alkyl bond with Ala306 (5.49 Å), Tyr311 (4.01 Å), Val314 (5.50 Å), Ala326 (4.62 Å, 5.06 Å) to stabilize within the binding pocket (**Figure 8B**). The highest binding affinity toward Luteolin was showed by OsUSP32 by forming one pi-pi-T-shaped bond with Phe374 (4.32 Å), two alkyl bond, one pi-alkyl bond, one pi-sigma with Ala306 (4.53 Å, 4.24 Å), Ala326 (5.35 Å), Val314 (4.43 Å), respectively (**Figure 8A**).

DISCUSSION

Universal stress proteins have been shown to be induced under various abiotic stresses. To date, a diverse number of *USP* gene family have been identified from different plant species including, *Arabidopsis* (Bhuria et al., 2019), *Brassica napus, Triticum aestivum, Brassica rapa, Solanum lycopersicum, Solanum tuberosum, Oryza sativa japonica, Vitis vinifera, Zea mays* (Chi et al., 2019), *Malus sieversii* (Ledeb.) (Yang et al., 2019), Barley

FIGURE 7 | Functional validation of selective *OsUSP* genes expression in response to various abiotic stresses and hormonal treatments by qRT-PCR. Expression profiles of nine *OsUSP* genes were performed in response to salinity, drought, cold, heat, dehydration, and oxidative stresses, and ABA, SA and GA3 treatments in BRRI-53 rice variety. A scale showing Log_2-fold change of each gene under each condition was presented at the left. The results were represented by the mean value of fold change ± standard deviation using rice *eEF1* gene as a reference. Statistical significance was determined using two-tailed paired Student's *t*-test and are represented with *, **, *** for a *P*-value < 0.05, <0.01, and <0.001; respectively.
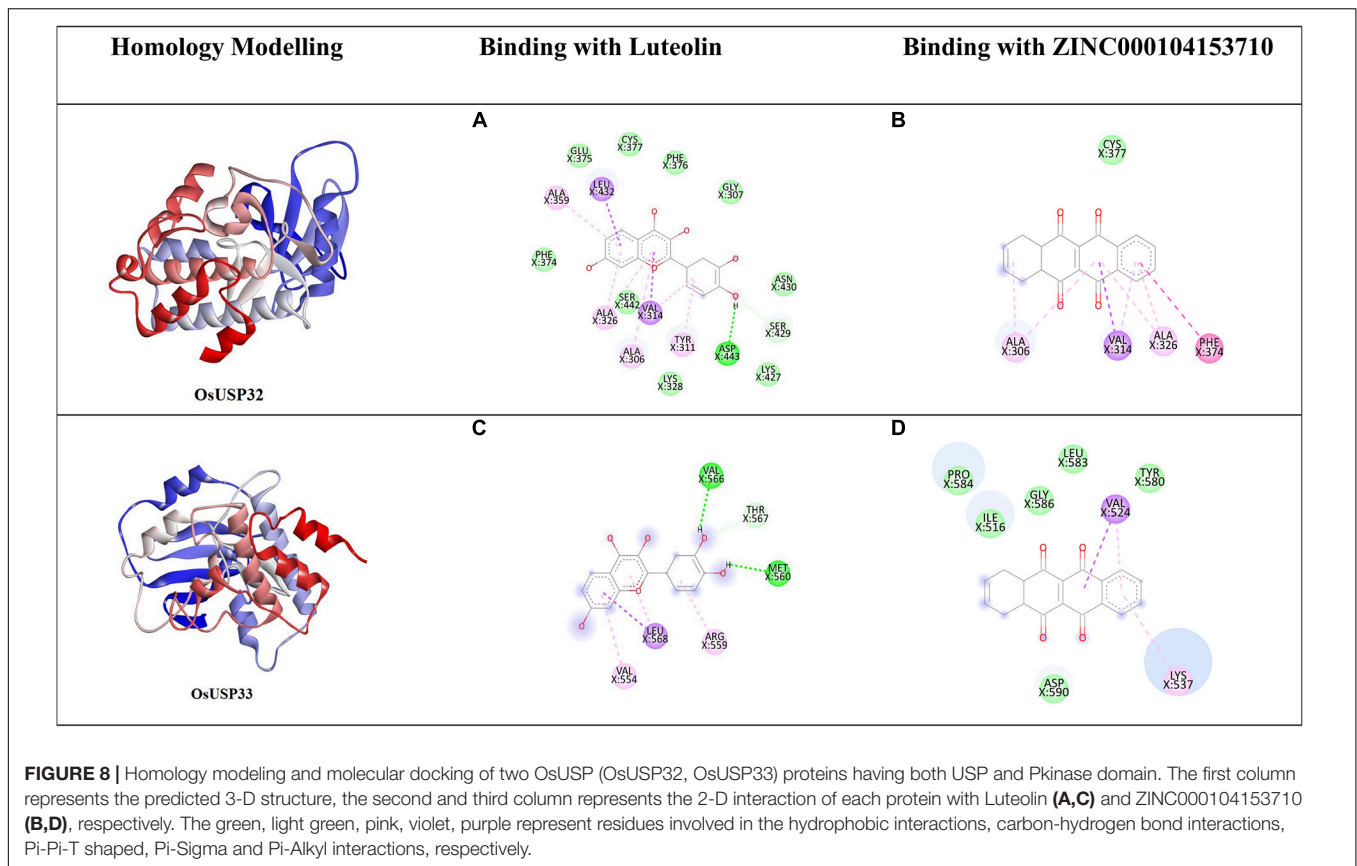
(Li et al., 2010), Pigeon pea (Sinha et al., 2016) etc. Usually, plant genomes harbor around twenty to fifty genes (Gou et al., 2020). However, more than a hundred members have been reported in the genome of *Brassica napus* and *Triticum aestivum* (Li et al., 2010; Chi et al., 2019). In the present study, a total of 44 *OsUSP* genes (**Table 1**) were identified in the rice genome.

The identified *USP* genes are spread throughout 11 out of 12 rice chromosomes (**Figure 3A**). The number of genes is slightly higher than previously reported (Bhowmick, 2019; Chi et al., 2019). Investigation into the structural properties of the identified genes and their corresponding protein showed that domain architecture played an important part in this. Analysis of synteny between rice *USP* genes showed that 16 out of 44 genes had traces of segmental duplication whereas two genes on chromosome 12 were likely derived from a tandem duplication event (**Figure 3B**). Thus, it appears that both segmental and tandem duplication

played a part in the expansion of the *USP* gene family in rice. Further investigation into the syntenic relationship between rice *USP* genes and the genome of *Arabidopsis thaliana* and *Zea mays* (maize) revealed the presence of a higher number of orthologs in the monocot maize genome compared to the dicot *Arabidopsis thaliana* (**Figure 3B**). This indicates greater diversification of *USP* genes between the two angiosperms lineages – monocots and dicots compared to divergence within each lineage. In this regard, our study suggests that both duplication and diversification have played a major role in the evolution of the *USP* gene family in plants.

Analyses of the OsUSPs phylogeny revealed two clusters that finely coincides with the domain architecture of the proteins. The cluster (green) that groups with *E. coli* USPs likely represents those members that originated earlier in the evolutionary history of the *USP* gene family. In contrast, members from

FIGURE 8 | Homology modeling and molecular docking of two OsUSP (OsUSP32, OsUSP33) proteins having both USP and Pkinase domain. The first column represents the predicted 3-D structure, the second and third column represents the 2-D interaction of each protein with Luteolin **(A,C)** and ZINC000104153710 **(B,D)**, respectively. The green, light green, pink, violet, purple represent residues involved in the hydrophobic interactions, carbon-hydrogen bond interactions, Pi-Pi-T shaped, Pi-Sigma and Pi-Alkyl interactions, respectively.

the other cluster (yellow) showed longer peptide lengths and the presence of a functional kinase domain. However, there were a few exceptions here, like *OsUSP10* and *OsUSP41*. Two *USP* genes could have undergone mutations that have rendered an existing kinase domain non-functional, or they could be on the verge of developing a kinase domain. The latter case is less likely, as these two USPs do not group closer to the green cluster which supposedly originated first. The existence of diverse domains in USP members in the yellow cluster signifies the functional diversity of USP members in plants. These proteins may have evolved to serve varied functions in protecting the plant from diverse stressors. Thus, it is imperative to accurately identify and further explore the diversity of these genes.

Molecular markers can be used as a selection tool for investigating genetic diversity, transferability of genes for genome editing purposes. In our study, we identified 19 SSR markers based on their high reproducibility, polymorphic genetic information, and hypervariable nature. As they produce high allelic variations among very closely related varieties, these markers can be used for the identification of specific *USP* members in future. In concordance with previous studies in Viridiplantae, we also found trinucleotide type as the dominant motif (Isokpehi et al., 2011).

Looking at the expression pattern of *OsUSP* genes we found that several members showed modulation under different abiotic stress conditions (**Figure 6A**). A number of these genes exhibited

strong upregulation when exposed to these stressors. These findings fit well with previous reports of the stress-specific regulation of *USP* genes (Loukehaich et al., 2012; Jung et al., 2015; Melencion et al., 2017). Moreover, it has been demonstrated that overexpression of certain *USP* genes lead to enhanced tolerance to heat and osmotic stress. In relation to this, candidate *OsUSP* genes may also have similar functional roles in stress tolerance. On top of this, we found the presence of multiple stress-responsive regulatory elements upstream of *OsUSP* genes. These include – ABRE (ABA response), ARE (anoxia), LTR (low temperature), MBS (drought), TC-rich repeats (stress and defense), HSE (heat shock), etc. A previous report by Bhuria et al. (2016) showed that *AtUSP* promoter is highly inducible by multiple abiotic stressors and was able to produce multi-stress tolerance. In another study in cotton plants, the promoter of a *USP* gene was found to be responsive to dehydration, ABA, salt, heavy metals, and gibberellic acid (Zahur et al., 2009). Moreover, Wild tomato *USP* genes were induced under ABA, ethylene, drought, salt, heat, wounding, oxidative, and cold stress (Loukehaich et al., 2012). Similarly, the *SbUSP* gene from *Salicornia brachiata* also showed expression under salt, drought, cold, and heat stress (Udawat et al., 2014). Overall, these provide extensive evidence for the differential modulation of rice *USP* genes under abiotic stress conditions and that their promoters have the capacity to induce stress resilience, making them good candidates to drive multiple stress-responsive expression of transgenes in genetically engineered plants.

TABLE 2 | Predicted binding affinity (kcal/mol) of selected OsUSP proteins against known USP and kinase inhibitors.

OsUSP	Binding affinity toward USP inhibitors (kcal/mol)			Binding affinity toward kinase inhibitors (kcal/mol)				Center grid box (points in X, Y, Z-axis)	Dimensions (points in X, Y, Z-axis)
	ZINC00010 4153710	ZINC00000 4268284	ZINC000000 217308	Fisetin	Luteolin	Myricetin	Quercetin		
OsUSP2	−7.6	−0.6.6	−7.3	Not performed				12.9957 × 40.7169 × 20.2698	42.9413 × 54.5348 × 38.3895
OsUSP3	−6.9	−6.5	−7.1	Not performed				12.2149 × 14.507 × 30.1618	51.9887 × 39.8425 × 42.0554
OsUSP6	−9.0	−7.1	−8.8	Not performed				7.8863 × 3.4420 × 36.2376	57.7571 × 53.4284 × 34.9141
OsUSP12	−8.7	−7.2	−8.1	Not performed				25.4699 × 7.9124 × 21.9120	56.2974 × 44.9545 × 47.3260
OsUSP14	−8.4	−8.4	−8.6	Not performed				19.775 × 16.9896 × 5.4766	50.7828 × 52.9327 × 42.3257
OsUSP22	−7.8	−7.5	−7.5	Not performed				56.0183 × 11.0732 × 11.3009	42.7089 × 59.3073 × 69.2370
OsUSP32	−9.6	−8.5	−8.9	−8.3	−9.2	−8.1	−8.1	44.0062 × 32.1862 × 65.9105	60.2972 × 44.9794 × 57.9650
OsUSP33	−8.5	−7.4	−8.2	−7.7	−8.2	−7.6	−7.7	15.8940 × 15.2491 × 33.7449	48.3649 × 60.8299 × 58.4618
OsUSP36	−8.3	−8.6	−7.7	Not performed				22.4725 × 73.5770 × 1.2069	46.4691 × 47.2079 × 54.9569

In coherence with this, during *in silico* expression profiling of *OsUSP* genes, we also found that expression of most rice *USP* genes was closely associated with heat, salinity, drought, dehydration, and submergence. Next, we used qRT-PCR to correlate the stress-specific function of the selected *USP* genes under multiple abiotic stresses. These led to the confirmation of the modulation of these rice *USP* genes under varying environmental stressors. For example, *OsUSP3* showed upregulation under all stress-conditions except cold which makes it a very suitable for downstream wet-lab validation (**Figure 7**). On the contrary, *OsUSP22* was mostly downregulated in the abiotic stressors except ABA and SA, indicating a strong modulation under stress. Interestingly, we observed that only *OsUSP3* and *OsUSP12* have exhibit elevated expression under oxidative stress. In addition, the response of each *USP* to the variety of stressors varied widely, indicating the possibility of diverse modes of transcriptional regulation for members of this gene family in rice. Uncovering these exact natures of such regulatory modules and their functional consequences can be of great importance for plant stress-oriented research.

To further bolster our understanding the rice *USP* gene family, we focused on analyzing the protein products of these genes. Proper functioning of proteins depends greatly on the post-translational modifications (PTMs) of proteins, glycosylation and phosphorylation being major two. In this regard, the glycosylation process plays a major role in protein folding, stability and signal transduction (Arey, 2012), whereas phosphorylation plays a critical role in the activation or deactivation of proteins by altering structural conformation (Cohen, 2002) and plays a significant role in the signaling pathways (Cutillas, 2017), metabolism and in combating various stress conditions (Zhang et al., 2020). In response to abiotic stresses in crop plants, PTM of proteins have been reported in response to abiotic stresses in crop plants and is suggested to be involved in various biological processes under natural conditions (Wu et al., 2016). Previous studies on protein PTM in soybean and rice under stresses showed that glycosylation occurs in response to flood and cold (Wu et al., 2016). Here, we identified 68 predicted glycosylation sites, among which 19 sites showed the highest chance of glycosylation. This confirms the ability of OsUSP members to perform glycosylation mediated secondary structure modification that may regulate the protein functions in stress condition. Furthermore, various studies suggested that different kinases are activated following that their activation loop residues are phosphorylated (Ravala et al., 2015) and autophosphorylation of UspA protein of *E. coli* also occur in response to stasis (Freestone et al., 1997). Moreover, phosphorylation mediated PTM was found prominent in combating multiple abiotic stressors in rice, maize, wheat, and chickpea (Wu et al., 2016). Thus, the predicted phosphorylation sites with position and specific kinases from the present work could be used for further confirmation of the role of phosphorylation in the function of different OsUSPs by using various *in vitro* techniques.

Moreover, a deep understanding of the binding affinity and binding site residues of inhibitors to OsUSPs is needed for the improvement of abiotic stress-resistant crops. In this present

study, we have performed blind docking (Hetényi and van der Spoel, 2002) method for exploration of possible binding sites and binding affinity of our proteins and toward some known USP and kinase inhibitors. In concordance with the previous study on UspA inhibitor in *E. coli* (Bandyopadhyay et al., 2021), ZINC000104153710 showed the best docking result, thus suggesting having an inhibitory effect on OsUSP activity. Furthermore, studies have found flavonoids acting as an inhibitor toward kinases (Cassidy and Setzer, 2010). Here, Luteolin showed the best result when docked against kinase domain containing OsUSPs. This strongly suggests that Luteolin may hinder kinase activity upon binding. This will facilitate further simulation, investigation regarding the interaction of other inhibitors to these binding sites and identification of some novel inhibitors that may inhibit the activity of USP proteins.

CONCLUSION

Overall, this study covers a comprehensive genome-wide scale investigation of *USP* gene family members in a staple crop, rice. In total, we were able to identify 44 *USP* genes in the monocot genome, a number slightly higher than previous reports. The key focus of this investigation was to perform an in-depth exploration of the structural and functional properties of the identified genes. This led us to discover the functional diversity of rice *USP* genes and their multi-stress regulatory nature. Moreover, using qRT-PCR, we were able validate this multi-stress responsive nature of promising *OsUSP* genes. However, further research is required to identify the functional contribution of these genes by generating overexpressing or knockout lines of *USP* genes alone or in combination. With the aid of modern biotechnological tools, the selected *OsUSP* genes could be targeted for improving not only rice but also other important crop species for developing multi-stress resilience plants.

REFERENCES

Arey, B. J. (2012). "The role of glycosylation in receptor signaling," in *Glycosylation*, ed. S. Petrescu (London: IntechOpen), doi: 10.5772/50262

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335

Bandyopadhyay, D., Singh, G., Mukherjee, M., and Akhter, Y. (2021). Computational approach towards the design of novel inhibitors against Universal stress protein A to combat multidrug resistant uropathogenic *Escherichia coli. J. Mol. Struc.* 1238:130379. doi: 10.1016/j.molstruc.2021.130379

Barnabás, B., Jäger, K., and Fehér, A. (2008). The effect of drought and heat stress on reproductive processes in cereals. *Plant Cell Environ.* 31, 11–38. doi: 10.1111/j.1365-3040.2007.01727.x

Basso, M. F., Arraes, F. B. M., Grossi-de-Sa, M., Moreira, V. J. V., Alves-Ferreira, M., and Grossi-de-Sa, M. F. (2020). Insights into genetic and molecular elements for transgenic crop development. *Front. Plant Sci.* 11:509. doi: 10.3389/fpls.2020.00509

Beier, S., Thiel, T., Münch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33, 2583–2585. doi: 10.1093/bioinformatics/btx198

Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., et al. (2015). The *Arabidopsis* information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis* 53, 474–485. doi: 10.1002/dvg.22877

Bhowmick, R. (2019). Comprehensive analysis of universal stress proteins and their promoter sequences in rice. *Int. J. Curr. Microbiol. App. Sci.* 8, 1279–1286. doi: 10.20546/ijcmas.2019.807.152

Bhuria, M., Goel, P., Kumar, S., and Singh, A. K. (2016). The promoter of AtUSP Is Co-regulated by phytohormones and abiotic stresses in *Arabidopsis thaliana. Front. Plant Sci.* 7:1957. doi: 10.3389/fpls.2016.01957

Bhuria, M., Goel, P., Kumar, S., and Singh, A. K. (2019). Genome-wide identification and expression profiling of genes encoding universal stress proteins (USP) identify multi-stress responsive USP genes in *Arabidopsis thaliana. Plant Physiol. Rep.* 24, 434–445. doi: 10.1007/s40502-019-00468-6

BIOVIA (2021). *Dassault Systèmes. in Discovery Studio Visualizer. v.21.10.20298.* San Diego, CA: Dassault Systèmes.

Blom, N., Gammeltoft, S., and Brunak, S. (1999). Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Bio.* 294, 1351–1362. doi: 10.1006/jmbi.1999.3310

Blom, N., Sicheritz-Ponten, T., Gupta, R., Gammeltoft, S., and Brunak, S. (2004). Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* 4, 1633–1649. doi: 10.1002/pmic.200300771

Cassidy, C. E., and Setzer, W. N. (2010). Cancer-relevant biochemical targets of cytotoxic Lonchocarpus flavonoids: a molecular docking analysis. *J. Mol. Mod.* 16, 311–326. doi: 10.1007/s00894-009-0547-5

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Chi, Y. H., Koo, S. S., Oh, H. T., Lee, E. S., Park, J. H., Phan, K. A. T., et al. (2019). The physiological functions of universal stress proteins and their molecular mechanism to protect plants from environmental stresses. *Front. Plant Sci.* 10:750. doi: 10.3389/fpls.2019.00750

Cohen, P. (2002). The origins of protein phosphorylation. *Nat. Cell Biol.* 4, E127–E130. doi: 10.1038/ncb0502-e127

Cutillas, P. R. (2017). Targeted in-depth quantification of signaling using label-free mass spectrometry. *Methods Enzymol.* 585, 245–268. doi: 10.1016/bs.mie.2016.09.021

Dallakyan, S., and Olson, A. J. (2015). Small-Molecule library screening by docking with PyRx. *Chem. Biol.* 1263, 243–250. doi: 10.1007/978-1-4939-2269-7_19

Emanuelsson, O., Nielsen, H., and Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* 8, 978–984. doi: 10.1110/ps.8.5.978

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223

Freestone, P., Nyström, T., Trinei, M., and Norris, V. (1997). The universal stress protein, UspA, of *Escherichia coli* is phosphorylated in response to stasis. *J. Mol. Bio.* 274, 318–324. doi: 10.1006/jmbi.1997.1397

Gabler, F., Nam, S. Z., Till, S., Mirdita, M., Steinegger, M., Söding, J., et al. (2020). Protein sequence analysis using the MPI bioinformatics toolkit. *Curr. Protoc. Bioinfor.* 72:e108. doi: 10.1002/cpbi.108

Gaut, B. S., Morton, B. R., McCaig, B. C., and Clegg, M. T. (1996). Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene Adh parallel rate differences at the plastid gene rbcL. *Proc. Natl. Acad. Sci. U.S.A.* 93, 10274–10279. doi: 10.1073/pnas.93.19.10274

Gou, L., Zhuo, C., Lu, S., and Guo, Z. (2020). A Universal Stress Protein from Medicago falcata (MfUSP1) confers multiple stress tolerance by regulating antioxidant defense and proline accumulation. *Environ. Exp. Bot.* 178:104168. doi: 10.1016/j.envexpbot.2020.104168

Gupta, R., and Brunak, S. (2002). Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac. Symp. Biocomput.* 2002, 310–322.

Gutiérrez-Beltrán, E., Personat, J. M., Torre, F., and Pozo, O. D. (2017). A universal stress protein involved in oxidative stress is a phosphorylation target for protein kinase CIPK6. *Plant Phy.* 173, 836–852. doi: 10.1104/pp.16.00949

Hasan, M. S., Singh, V., Islam, S., Islam, M. S., Ahsan, R., Kaundal, A., et al. (2021). Genome-wide identification and expression profiling of glutathione S-transferase family under multiple abiotic and biotic stresses in Medicago truncatula L. *PLoS One* 16:e0247170. doi: 10.1371/journal.pone.0247170

Hetényi, C., and van der Spoel, D. (2002). Efficient docking of peptides to proteins without prior knowledge of the binding site. *Prot. Sci.* 11, 1729–1737. doi: 10.1110/ps.0202302

Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: protein localization predictor. *Nucl. Acids Res.* 35, W585–W587. doi: 10.1093/nar/gkm259

Howe, K. L., Contreras-Moreira, B., De Silva, N., Maslen, G., Akanni, W., Allen, J., et al. (2020). Ensembl Genomes 2020—enabling non-vertebrate genomic research. *Nucleic Acids Res.* 48, D689–D695. doi: 10.1093/nar/gkz890

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., et al. (2008). Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv. Bioinform.* 2008, 420747. doi: 10.1155/2008/420747

Islam, T., Manna, M., and Kaul, T. (2015). Genome-Wide dissection of *Arabidopsis* and rice for the identification and expression analysis of glutathione peroxidases reveals their stress-specific and overlapping response patterns. *Plant Mol. Biol. Rep.* 33, 1413–1427. doi: 10.1007/s11105-014-0846-6

Isokpehi, R. D., Simmons, S. S., Cohly, H. H. P., Ekunwe, S. I. N., Begonia, G. B., and Ayensu, W. K. (2011). Identification of drought-responsive universal stress proteins in Viridiplantae. *Bioinform. Biol. Insights* 5, 41–58. doi: 10.4137/BBI.S6061

Jung, Y. J., Melencion, S. M. B., Lee, E. S., Park, J. H., Alinapon, C. V., Oh, H. T., et al. (2015). Universal stress protein exhibits a redox-dependent chaperone function in *Arabidopsis* and enhances plant tolerance to heat shock oxidative stress. *Front. Plant Sci.* 6:1141. doi: 10.3389/fpls.2015.01141

Katoh, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 20, 1160–1166. doi: 10.1093/bib/bbx108

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., et al. (2021). PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* 49, D1388–D1395. doi: 10.1093/nar/gkaa971

Lee, T. H., Tang, H., Wang, X., and Paterson, A. H. (2012). PGDD: a database of gene and genome duplication in plants. *Nucleic Acids Res.* 41(D1), D1152–D1158. doi: 10.1093/nar/gks1104

Lescot, M. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30, 325–327. doi: 10.1093/nar/30.1.325

Lesk, C., Rowhani, P., and Ramankutty, N. (2016). Influence of extreme weather disasters on global crop production. *Nature* 529, 84–87. doi: 10.1038/nature16467

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

Li, W. T., Wei, Y. M., Wang, J. R., Liu, C. J., Lan, X. J., Jiang, Q. T., et al. (2010). Identification, localization, and characterization of putative USP genes in barley. *Theor. Appl. Genet.* 121, 907–917. doi: 10.1007/s00122-010-1359-9

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2 - \Delta\Delta CT$ method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Loukehaich, R., Wang, T., Ouyang, B., Ziaf, K., Li, H., Zhang, J., et al. (2012). SpUSP, an annexin-interacting universal stress protein, enhances drought tolerance in tomato. *J. Exp. Bot.* 63, 5593–5606. doi: 10.1093/jxb/ers220

Maqbool, A., Zahur, M., Husnain, T., and Riazuddin, S. (2009). GUSP1 and GUSP2, two drought-responsive genes in Gossypium arboreum have homology to universal stress proteins. *Plant Mol. Biol. Rep.* 27, 107–114. doi: 10.1007/s11105-008-0049-0

Melencion, S. M. B., Chi, Y. H., Pham, T. T., Paeng, S. K., Wi, S. D., Lee, C., et al. (2017). RNA chaperone function of a universal stress protein in *Arabidopsis* confers enhanced cold stress tolerance in plants. *Int. J. Mol. Sci.* 18:2546. doi: 10.3390/ijms18122546

Merkouropoulos, G., and Tsaftaris, A. S. (2013). Differential expression of Gossypium hirsutum USP-related genes, GhUSP1 and GhUSP2, during development and upon salt stress. *Plant Mol. Biol. Rep.* 31, 1539–1547. doi: 10.1007/s11105-013-0630-z

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). Corrigendum to: IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic Era. *Mol. Bio. Evol.* 37, 2461–2461. doi: 10.1093/molbev/msaa131

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2020). Pfam: the protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi: 10.1093/nar/gkaa913

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., et al. (2009). AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30, 2785–2791. doi: 10.1002/jcc.21256

Munns, R. (2002). Comparative physiology of salt and water stress: comparative physiology of salt and water stress. *Plant, Cell Environ.* 25, 239–250. doi: 10.1046/j.0016-8025.2001.00808.x

Nyström, T., and Neidhardt, F. C. (1992). Cloning, mapping and nucleotide sequencing of a gene encoding a universal stress protein in *Escherichia coli*. *Mol. Microbiol.* 6, 3187–3198. doi: 10.1111/j.1365-2958.1992.tb01774.x

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011). Open Babel: an open chemical toolbox. *J. Cheminform.* 3:33. doi: 10.1186/1758-2946-3-33

Ravala, S. K., Singh, S., Yadav, G. S., Kumar, S., Karthikeyan, S., and Chakraborti, P. K. (2015). Evidence that phosphorylation of threonine in the GT motif triggers activation of PknA, a eukaryotic-type serine/threonine kinase from Mycobacterium tuberculosis. *FEBS J.* 282, 1419–1431. doi: 10.1111/febs.13230

Sakai, H., Lee, S. S., Tanaka, T., Numa, H., Kim, J., Kawahara, Y., et al. (2013). Rice annotation project database (RAP-DB): An integrative and interactive database for rice genomics. *Plant Cell Physiol.* 54:e6. doi: 10.1093/pcp/pcs183

Sauter, M. (2002). The novel ethylene-regulated gene OsUsp1 from rice encodes a member of a plant protein family related to prokaryotic universal stress proteins. *J. Exp. Bot.* 53, 2325–2331. doi: 10.1093/jxb/erf096

Shinozaki, K., and Yamaguchi-Shinozaki, K. (2007). Gene networks involved in drought stress response and tolerance. *J. Exp. Bot.* 58, 221–227. doi: 10.1093/jxb/erl164

Sinha, P., Pazhamala, L. T., Singh, V. K., Saxena, R. K., Krishnamurthy, L., Azam, S., et al. (2016). Identification and validation of selected universal stress protein domain containing drought-responsive genes in pigeon pea (*Cajanus cajan* L.). *Front. Plant Sci.* 6:1065. doi: 10.3389/fpls.2015.01065

Smith, S. A., and Dunn, C. W. (2008). Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* 24, 715–716. doi: 10.1093/bioinformatics/btm619

Song, S., Tian, D., Zhang, Z., Hu, S., and Yu, J. (2019). Rice genomics: over the past two decades and into the future. *Geno. Prot. Bio.* 16, 397–404. doi: 10.1016/j.gpb.2019.01.001

Sousa, M. C., and McKay, D. B. (2001). Structure of the universal stress protein of *Haemophilus influenzae*. *Structure* 9, 1135–1141. doi: 10.1016/s0969-2126(01)00680-3

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., et al. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 45, W122–W129. doi: 10.1093/nar/gkx382

Trott, O., and Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461. doi: 10.1002/jcc.21334

Udawat, P., Jha, R. K., Sinha, D., Mishra, A., and Jha, B. (2016). Overexpression of a cytosolic abiotic stress responsive universal stress protein (SbUSP) mitigates salt and osmotic stress in transgenic tobacco plants. *Front. Plant Sci.* 7:518. doi: 10.3389/fpls.2016.00518

Udawat, P., Mishra, A., and Jha, B. (2014). Heterologous expression of an uncharacterized universal stress protein gene (SbUSP) from the extreme halophyte, *Salicornia brachiata*, which confers salt and osmotic tolerance to E. coli. *Gene* 536, 163–170. doi: 10.1016/j.gene.2013.11.020

UniProt Consortium (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49, D480–D489. doi: 10.1093/nar/gkaa1100

Vollmer, A. C., and Bark, S. J. (2018). Twenty-Five years of investigating the universal stress protein: function, structure, and applications. *Adv. App. Microbiol.* 102, 1–36. doi: 10.1016/bs.aambs.2017.10.001

Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40:e49. doi: 10.1093/nar/gkr1293

Wu, X., Gong, F., Cao, D., Hu, X., and Wang, W. (2016). Advances in crop proteomics: PTMs of proteins under abiotic stress. *Proteomics* 16, 847–865. doi: 10.1002/pmic.201500301

Yang, M., Che, S., Zhang, Y., Wang, H., Wei, T., Yan, G., et al. (2019). Universal stress protein in Malus sieversii confers enhanced drought tolerance. *J. Plant Res.* 132, 825–837. doi: 10.1007/s10265-019-01133-7

Yu, C. S., Chen, Y. C., Lu, C. H., and Hwang, J. K. (2006). Prediction of protein subcellular localization. *Proteins* 64, 643–651. doi: 10.1002/prot.21018

Zahur, M., Maqbool, A., Ifran, M., Barozai, M. Y., Rashid, B., Riazuddin, S., et al. (2009). Isolation and functional analysis of cotton universal stress protein promoter in response to phytohormones and abiotic stresses. *Mol. Biol.* 43, 578–585. doi: 10.1134/S0026893309040086

Zarembinski, T. I., Hung, L. W., Mueller-Dieckmann, H. J., Kim, K. K., Yokota, H., Kim, R., et al. (1998). Structure-based assignment of the biochemical function hypothetical protein: a test case of structural genomics. *Proc. Natl. Acad. Sci. U.S.A.* 95, 15189–15193. doi: 10.1073/pnas.95.26.15189

Zhang, X., Zhuang, L., Liu, Y., Yang, Z., and Huang, B. (2020). Protein phosphorylation associated with drought priming-enhanced heat tolerance in a temperate grass species. *Hortic. Res.* 7:207. doi: 10.1038/s41438-020-00440-8

Zimmermann, L., Stephens, A., Nam, S. Z., Rau, D., Kübler, J., Lozajic, M., et al. (2018). A completely reimplemented MPI bioinformatics toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* 430, 2237–2243.

Structural and Evolutionary Analyses of PR-4 SUGARWINs Points to a Different Pattern of Protein Function

Lorhenn Bryanda Lemes Maia[1], Humberto D'Muniz Pereira[2], Richard Charles Garratt[2], José Brandão-Neto[3], Flavio Henrique-Silva[4], Danyelle Toyama[4], Renata O. Dias[5], José Fernando Ruggiero Bachega[6,7], Julia Vasconcellos Peixoto[7] and Marcio C. Silva-Filho[1,8]*

[1] Programa de Pós-Graduação em Biotecnologia, Universidade Católica Dom Bosco, Campo Grande, Brazil, [2] Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, Brazil, [3] Diamond Light Source, Harwell Science and Innovation Campus Didcot, Harwell, United Kingdom, [4] Departamento de Genética e Evolução, Universidade Federal de São Carlos, São Carlos, Brazi, [5] Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, Brazil, [6] Departamento de Farmacociências, Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, Brazil, [7] Programa de Pós-Graduação de Biologia Celular e Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, [8] Departamento de Genética, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, Brazil

SUGARWINs are PR-4 proteins associated with sugarcane defense against phytopathogens. Their expression is induced in response to damage by *Diatraea saccharalis* larvae. These proteins play an important role in plant defense, in particular against fungal pathogens, such as *Colletothricum falcatum* (Went) and *Fusarium verticillioides*. The pathogenesis-related protein-4 (PR-4) family is a group of proteins equipped with a BARWIN domain, which may be associated with a chitin-binding domain also known as the hevein-like domain. Several PR-4 proteins exhibit both chitinase and RNase activity, with the latter being associated with the presence of two histidine residues H11 and H113 (BARWIN) [H44 and H146, SUGARWINs] in the BARWIN-like domain. In sugarcane, similar to other PR-4 proteins, SUGARWIN1 exhibits ribonuclease, chitosanase and chitinase activities, whereas SUGARWIN2 only exhibits chitosanase activity. In order to decipher the structural determinants involved in this diverse range of enzyme specificities, we determined the 3-D structure of SUGARWIN2, at 1.55Å by X-ray diffraction. This is the first structure of a PR-4 protein where the first histidine has been replaced by asparagine and was subsequently used to build a homology model for SUGARWIN1. Molecular dynamics simulations of both proteins revealed the presence of a flexible loop only in SUGARWIN1 and we postulate that this, together with the presence of the catalytic histidine at position 42, renders it competent as a ribonuclease. The more electropositive surface potential of SUGARWIN1 would also be expected to favor complex formation with RNA. A phylogenetic analysis of PR-4 proteins obtained from 106 Embryophyta genomes showed that both catalytic histidines are widespread among them with few replacements in these amino acid positions during the gene family evolutionary history. We observe that the H11 replacement by N11 is also present in two other sugarcane

PR-4 proteins: SUGARWIN3 and SUGARWIN4. We propose that RNase activity was present in the first Embryophyta PR-4 proteins but was recently lost in members of this family during the course of evolution.

INTRODUCTION

Plant responses to insects and pathogens are complex and modulate the expression of a large number of genes, many of which are believed to have a direct role in plant defense (Xu et al., 1994; Agrawal et al., 2003; Banno et al., 2005; Franco et al., 2017). Pathogen recognition by plants activates the host defense response resulting in the accumulation of pathogenesis-related proteins (PR proteins) (Pieterse and van Loon, 1999). Several genes that encode wound-inducible proteins (WIN) have been identified in various plant species (Ryan, 1990; Christopher et al., 2004; Schilmiller and Howe, 2005; Medeiros et al., 2012).

The pathogenesis-related protein-4 (PR-4) family is a group of proteins composed of a BARWIN domain, which was first identified in the BARWIN protein from barley (Hejgaard et al., 1992; Svensson et al., 1992). Although the BARWIN domain can be found in association with a chitin-binding domain typical of lectins, known as a hevein-like domain (Broekaert et al., 1990), several proteins show the absence of this protein motif. PR-4 proteins composed only by BARWIN domains were identified in several plants, such as: tobacco (Friedrich et al., 1991), tomato (Linthorst et al., 1991), *Arabidopsis* (Potter et al., 1993), wheat (Caruso et al., 1999), *W. japonica* (Kiba et al., 2003), maize (Bravo et al., 2003), rice (Agrawal et al., 2003; Zhu et al., 2006) and *L. radiata* (Li et al., 2010). The PR-4 proteins are grouped into class I and class II based on the presence or absence of this chitin-binding domain (Neuhaus et al., 1996).

The PR-4 BARWIN homologs of several plant species have been associated with the plant responses to fungal infection and mechanical wounding (Linthorst et al., 1991; Hejgaard et al., 1992; Caruso et al., 1999, 2001; Agrawal et al., 2003; Bravo et al., 2003; Kiba et al., 2003; Zhu et al., 2006; Li et al., 2010; Medeiros et al., 2012; Bai et al., 2013; Franco et al., 2014, 2019; Menezes et al., 2014; Souza et al., 2017). PR-4 proteins have been classified as chitinases (Neuhaus et al., 1996; Van Loon and Van Strien, 1999); however, several studies have also reported RNase activity for BARWIN-like proteins (Bertini et al., 2009, 2012; Guevara-Morato et al., 2010; Bai et al., 2013; Huet et al., 2013; Franco et al., 2014; Menezes et al., 2014; Kim and Hwang, 2015). RNA-binding sites have been described for WHEATWIN1 (Bai et al., 2013) and CARWIN (Huet et al., 2013), showing two important histidine residues necessary for RNase activity, one at position 11 and another at position 113, numbered according to the mature BARWIN sequence, which may be used as the reference residue numbering in this text (Bai et al., 2013). Combined DNase and RNase antifungal activities were also observed for the *Capsicum chinense* PR-4 protein (Guevara-Morato et al., 2010) and the *Theobroma cacao* TcPR-4b protein (Menezes et al., 2014).

Our previous studies have identified two homologous genes to BARWIN in sugarcane: *SUGARWIN1* and *SUGARWIN2*

(Falco et al., 2001; Medeiros et al., 2012). SUGARWIN proteins (sugarcane wound-inducible proteins) are believed to be part of a defense mechanism against pathogenic fungi in sugarcane plants *via* a salicylic acid (SA)-independent and jasmonic acid (JA)- dependent pathway. These proteins are secreted to the apoplasts, are strongly upregulated in response to mechanical wounding, *Diatraea saccharalis* (Fabricius) attack, methyl jasmonate treatment and to a lesser extent to sugarcane fungus (Medeiros et al., 2012). Despite their high expression levels in response to *D. saccharalis* attack, these proteins have no effect on insect development and mortality. However, SUGARWINs have antimicrobial effects in pathogenic fungus causing changes in hyphae morphology and leading to cell death by apoptosis (Medeiros et al., 2012; Franco et al., 2014). We have recently shown that *Fusarium verticillioides*, a pathogenic fungus, modulates sugarcane plants and *Diatrea saccharalis* moths and larvae to increase its dissemination in the field (Franco et al., 2021). The antifungal activity of SUGARWIN proteins is likely due to its multiple enzyme properties, such as: chitosanase, RNase and chitinase activities. This variable enzyme specificity has been suggested to be a result of a divergent amino acid composition at the substrate- binding site of these two proteins (Franco et al., 2014).

In order to understand how these SUGARWINs acquired these divergent activities, here we analyze their evolutionary history and their three-dimensional structures, looking for a correlation between structural and functional properties. Furthermore, we also describe two new putative SUGARWIN genes: *SUGARWIN3* and *SUGARWIN4*, which may also lack the canonical RNase activity observed in other PR-4 proteins as both have asparagine instead of histidine in the amino acid position 11 (43—SUGARWIN3; 38—SUGARWIN4).

MATERIALS AND METHODS

Recombinant Expression and Purification of SUGARWIN2

The heterologous expression of SUGARWIN2 was performed in *Pichia pastoris* as previously described (Medeiros et al., 2012). Basically, a single colony of *P. pastoris* containing the pPICZα-SUG2, that contains the coding region for SUGARWIN2 was inoculated in 10 ml of BMGY medium [1% yeast extract, 2% peptone, 100 mM potassium phosphate buffer (pH 7.0), 1.34% YNB, 10^{-5}% biotin, and 1% glycerol], which was incubated at 30°C until reaching an optical density (OD) at 600 nm of about 5. This culture was used to inoculate 500 ml of BMGY and was grown to an OD of 4–5. The cells were harvested by centrifugation at 1,500 × *g* for 5 min, resuspended in

100 ml of BMGY with 0.5% methanol instead of glycerol, and incubated at 28°C. For induction, methanol was added to each sample every 24 h to maintain a final concentration of 0.75%. After 96 h, the cells were harvested by centrifugation at 1,500 × g for 5 min, and the supernatant was filtered with a 0.45 mm membrane (Millipore, Bedford, MA, United States). The recombinant protein in the supernatant was purified by immobilized metal affinity chromatography (IMAC) using Ni-NTA-agarose (Qiagen) pre-equilibrated with purification buffer (10 mM Tris–HCl, pH 8.0; 50 mM NaH2PO4; and 100 mM NaCl). After binding, the proteins were eluted with two-column volumes of purification buffer containing increasing imidazole concentrations (10, 25, 50, 75, 100, and 250 mM). The fractions containing the recombinant protein were pooled, dialyzed in phosphate-buffered saline (PBS, pH 7.4) (137 mM NaCl, 2.7 mM KCl, 10 mM Na2PO4 and 2 mM KH2PO4) and sterilized with a 0.22 mm filter (Millipore). The protein concentration was determined using a BCA protein assay kit (Pierce).

The purification of the protein was monitored using 12% SDS-PAGE (**Supplementary Figure 1**). The recombinant SUGARWIN2 sample was further purified in an AKTA Purifier (GE, Healthcare) with a gel filtration column (Superdex 75) in TRIS solution (20 mM; pH 7.4), with 200 mM of NaCl. The final purification of the recombinant protein SUGARWIN2 was monitored by Dynamic Light Scattering (DLS) (**Supplementary Figure 2**).

Crystallization, Data Collection, Structure Determination and Refinement of SUGARWIN2

Recombinant SUGARWIN2 from gel filtration experiment at 7 mg/mL were crystallized by the sitting drop vapor diffusion method using the Morpheus screening kit (Molecular Dimensions). 200 nL drops were setup employing a Crystal Gryphon robot (Art Robbins). The plate was incubated at 18°C. Several conditions resulted in bipyramidal crystals after 1 week. The crystals were flash-cooled in liquid nitrogen for data collection. The structure presented here is from a crystal grown in 0.1 M MES/imidazole pH 6.5, 10% w/v PEG 4000, 20% v/v glycerol 0.03 M sodium fluoride 0.03 M sodium bromide and 0.3 M sodium iodide.

X-ray diffraction data were collected on beamline I24 of the Diamond Light Source (DLS), United Kingdom, (λ = 0.96863 Å) using a PILATUS3 6 M detector. The diffraction data were indexed, integrated and scaled using the xia2 package (REF). The data were processed up to 1.55 Å resolution. The structure was solved by molecular replacement using Phaser (McCoy et al., 2007); and papaya BARWIN-like protein (CARWIN, PDB ID: 4JP7) as search model after modification using Chainsaw (Stein, 2008). The structure was refined using Phenix (Adams et al., 2011) and COOT (Emsley and Cowtan, 2004) was used for model building into σa-weighted 2Fo-Fc and Fo-Fc electron density maps. R and Rfree were monitored to evaluate the validity of the refinement protocol, and the stereochemistry of the model was assessed using Molprobity (Chen et al., 2010). The

coordinates and structure factors have been deposited in Protein Data Bank (PDB ID: 7KSN).

The data-collection, processing and refinement statistics are shown in **Table 1**.

Gene Sequences and Phylogenetic Analysis

Protein sequences homologous to SUGARWINs 1 and 2 were searched for in 106 Embryophyta genomes obtained from the NCBI RefSeq Genome database and in the genomes of *Saccharum spontaneum* (Zhang et al., 2018) and the Brazilian hybrid cultivar *Saccharum spp.* SP80-3280 (Souza et al., 2019), using the BLASTp algorithm (Altschul et al., 1990). For genes with multiple isoforms of splicing, only the longest one was retrieved for further analyses. Protein domains were predicted using the InterProScan software (Jones et al., 2014).

A Maximum Likelihood phylogenetic tree inference was predicted using the IQ-TREE software v.1.6.12 (Nguyen et al., 2015) employing the automatic amino acid substitution model selection. Branch support values were calculated using 100 bootstrap replications. The protein sequence alignment used in the phylogenetic tree analysis was predicted using

TABLE 1 | Data-collection, processing and refinement statistics.

	SUGARWIN2
Detector	Pilatus3 6M
Cell parameters (Å) a, b, c, α, β, γ	227.51, 227.51, 227.51, 90.0, 90.0, 90.0
Space group	*F*4132
Resolution (Å)	52.20–1.51 (1.55–1.55)
X ray source	DLS I24
λ (Å)	0.96863
Multiplicity	75.7 (72.2)
Rpim (%)	1.7 (42.1)
CC (1/2)	0.999 (0.791)
Completeness (%)	100.0 (100.0)
Reflections	5,976,527 (415,898)
Unique reflections	78,971 (5,764)
I/σ	23.2 (2.4)
Reflections used for refinement	78,910
R (%)	16.26
Rfree (%)	18.50
N° of protein atoms	1,848
N° of ligant atoms	0
B (Å²)	13.88
Coordinate error (Å)	0.13
Phase error (°)	16.52
Ramachandran plot	
Favored (%)	96.72
Allowed (%)	3.28
Outliers (%)	0.0
All-atom clashscore	4.23
RMSD from ideal geometry	
r.m.s. bond lengths (Å)	0.014
r.m.s. bond angles (°)	1.291
PDB ID	7KSN

MAFFT software v7.453 (Katoh and Standley, 2013) and filtered using trimAL v.1.4 with the -automated1 method (Capella-Gutierrez et al., 2009).

Molecular Modeling, Molecular Dynamics and Homology Modeling of SUGARWINs

All the SUGARWIN systems were prepared in AmberTool19 (Olsson et al., 2011) using the ff99SB force field for proteins and TIP3P for water molecules. The conversion of the original structures to the amber topology and coordinates, as well as the addition of water molecules and counter ions was performed with the *tleap* program. A buffer of at least 10 Å between protein and the periodic box wall was used. For the SUGARWIN2 model, the input coordinates were taken from the crystallographic structure presented in this work (PDB ID: 7KSN). For SUGARWIN1, SUGARWIN3 and SUGARWIN4 the coordinates were taken from those obtained from comparative models. Protonation states of the protein were assigned at pH 7.5 using the PROPKA3 (Olsson et al., 2011) web server.

The MD simulations were performed using NaMD, version 2.14 (Olsson et al., 2011). The time step and the temperature were set to 2.0 fs and 300 K, respectively. Simulations were performed in the isothermal-isobaric ensemble (NPT) and periodic boundary conditions were employed, with Van der Waals interactions computed using a 12.0 Å cut-off, and electrostatic contributions calculated *via* a particle mesh Ewald method using a grid with 1.0 Å spacing. Prior to MD, each of the systems was energy minimized with a steepest-descent energy algorithm. Altogether, MD simulations were performed for a duration of 100 ns after equilibration, with snapshots of the system being saved every 20 ps, giving 5,000 snapshots for subsequent analysis. All the trajectories were analyzed in Bio3D package (Grant et al., 2006).

The homology modeling was performed using Modeller v9.24 (Marti-Renom et al., 2000) employing the coordinates of the Sugarwin2 as model. 200 models were built for each SUGARWIN and the best model was selected by using internal Modeller energy score.

RESULTS

PR-4 Phylogenetic Proposition

A total of 436 PR-4 protein sequences were identified in our searches of Embryophyta genomes. These sequences were present in 101 out of 106 analyzed genomes, corresponding to 95.3% of the total. PR-4 genes were found to be absent from the five species of Asterales (3) and Lamiales (2), but were present in the form of multiple copies in 94 of the remaining genomes (**Supplementary Table 1**).

The BARWIN domain (Pfam ID: PF00967) was found in all protein sequences analyzed, whereas a "chitin recognition protein" domain (Pfam ID: PF00187) was found in 92, which corresponds to 21.1%. In addition, the chitin recognition domain was absent in all sequences from
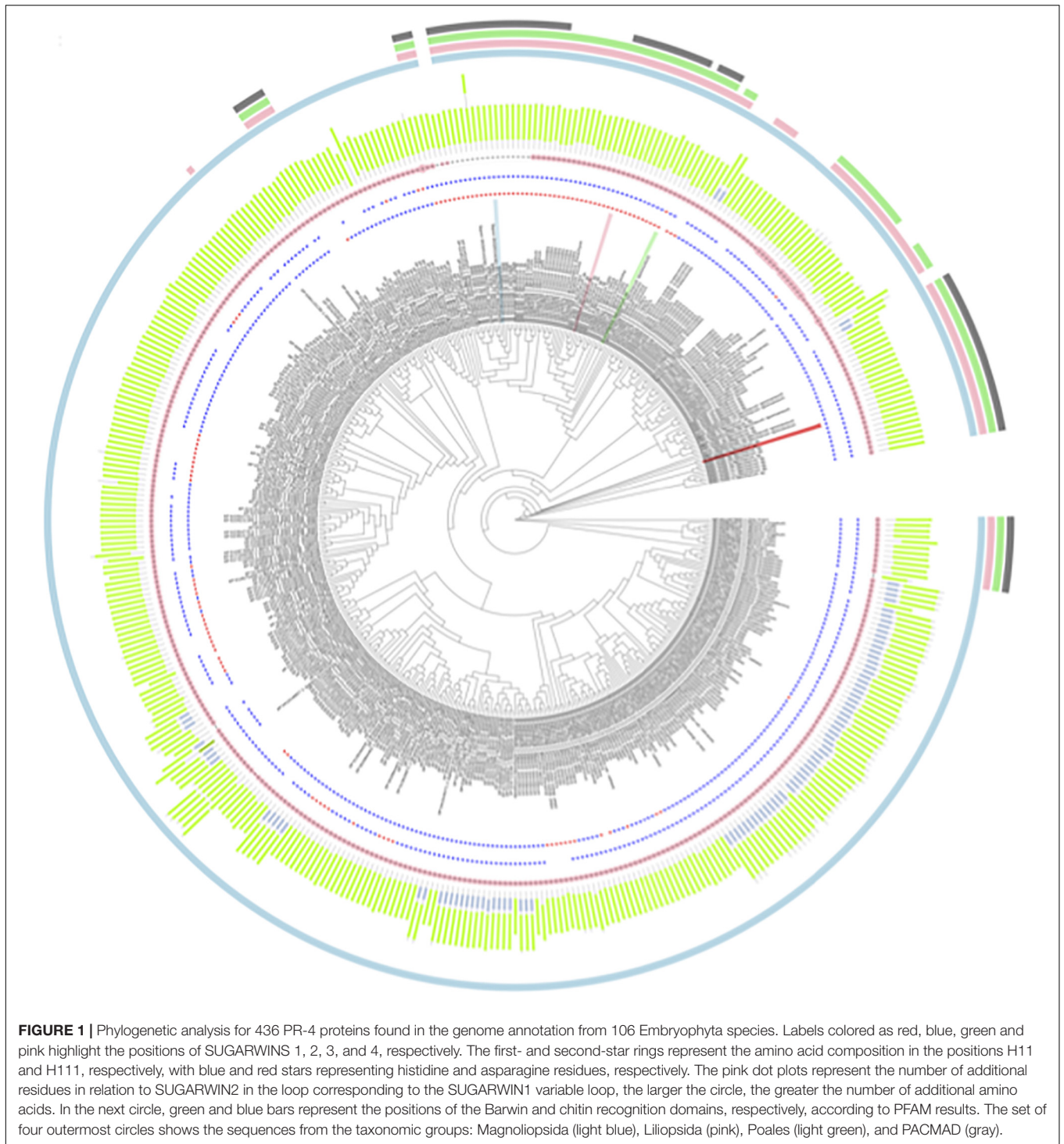
Liliopsida species. Among the 436 PR-4 protein sequences found in our searches, 312 and 101 of them had histidine or asparagine at position 11 [H42—SUGARWIN1; N37—SUGARWIN2; N43—SUGARWIN3; N38—SUGARWIN4], respectively, whereas 362 and 18 had histidine or asparagine at position 113 [H142—SUGARWIN1; H135—SUGARWIN2; N142—SUGARWIN3; H137—SUGARWIN4] (**Figure 1**). These two positions have been associated with the different catalytic activities attributed to PR-4 proteins and a total of 268 of the sequences examined (61.5%) had both histidine residues H11 and H113 in their catalytic site.

Crystallography and Molecular Modeling

PR-4 proteins lacking the His11-His113 pair are less common and do not have its crystal structure, therefore, we solved the crystal structure of SUGARWIN2. Analysis of the protein structure shows an asparagine residue at position 11, which is followed by a loop shorter by two residues when compared either with SUGARWIN1, either with the canonical structures of BARWIN and CARWIN.

SUGARWIN2 crystallizes in cubic space group $F4_132$ with two molecules per asymmetric unit. The structure was refined to R and R_{free} values of 16.26 and 18.50% respectively. The two molecules are in the same conformation with an RMSd of 0.18 Å on Cα positions. When compared with CARWIN from papaya, the RMSd is 0.45 Å, indicating a very well conserved 3D fold. SUGARWIN2 presents the classical BARWIN fold (CATH classification 2.40.40) consisting of a six-stranded β-barrel surrounded by 4 α-helices and several loops (**Figures 2, 3**). Three disulfide bonds (Cys30-Cys62, Cys51-Cys85, Cys65-Cys122) help to maintain the BARWIN domain stability.

Classical ribonucleases, such as RNase A, possess two catalytic histidine residues in their active sites. These residues are involved in a general acid-base mechanism leading to substrate hydrolysis (**Figure 3A**). The absence of RNase activity in SUGARWIN2 is believed to be due to the substitution of His11 by an asparagine. SUGARWIN1, BARWIN and CARWIN proteins have both histidines at the same positions and display ribonuclease activity. In both RNase A (PDB ID: 5RSA) and CARWIN (PDB ID: 4JP7) the distance between Cα's of the two histidines is about 10 Å. This is also the case in the inactive SUGARWIN2 structure reported here, albeit that one has been substituted by an asparagine residue. However, different from RNase A, where the histidine sidechains face one to another in a catalytically competent arrangement, CARWIN structure shows that these residues point to opposite directions. Similar observation is reported for the histidine/asparagine pair in SUGARWIN2. Taking all together, we suggest that SUGARWIN1 and other ribonuclease PR-4 proteins must provide a significant flexibility at the active site to assure the catalytic histidine properties for catalysis (see below). In addition, the histidine sidechains face one another in a catalytically competent arrangement, whereas in CARWIN they point in opposite directions and the same is true of the histidine/asparagine pair in SUGARWIN2 (**Figure 3**). Together, these observations suggest that in the case of SUGARWIN1 and other ribonuclease active PR-4 proteins, there must be significant flexibility to the active site region, in particular the

FIGURE 1 | Phylogenetic analysis for 436 PR-4 proteins found in the genome annotation from 106 Embryophyta species. Labels colored as red, blue, green and pink highlight the positions of SUGARWINS 1, 2, 3, and 4, respectively. The first- and second-star rings represent the amino acid composition in the positions H11 and H111, respectively, with blue and red stars representing histidine and asparagine residues, respectively. The pink dot plots represent the number of additional residues in relation to SUGARWIN2 in the loop corresponding to the SUGARWIN1 variable loop, the larger the circle, the greater the number of additional amino acids. In the next circle, green and blue bars represent the positions of the Barwin and chitin recognition domains, respectively, according to PFAM results. The set of four outermost circles shows the sequences from the taxonomic groups: Magnoliopsida (light blue), Liliopsida (pink), Poales (light green), and PACMAD (gray).

catalytic histidines, to re-orientate into positions appropriate for catalysis (see below).

According to our observations there is a variable loop following the first β-strand which is two residues shorter in SUGARWIN2 when compared with the ribonuclease-active SUGARWIN1, BARWIN and CARWIN. In the latter, the inserted residues are Lys-Val which form a type I β-turn, where

the valine carbonyl is hydrogen bonded to the catalytic His11 *via* a water molecule, thereby connecting the active site to the variable loop. The equivalent insertion in SUGARWIN1 is Asn-Ala and would be expected to behave similarly. In addition, for the SUGARWIN2 the shorter loop forms a type II β-turn (facilitated by the presence of Gly24) and there is no connectivity to the asparagine, which replaces His11 at the active site. Instead,
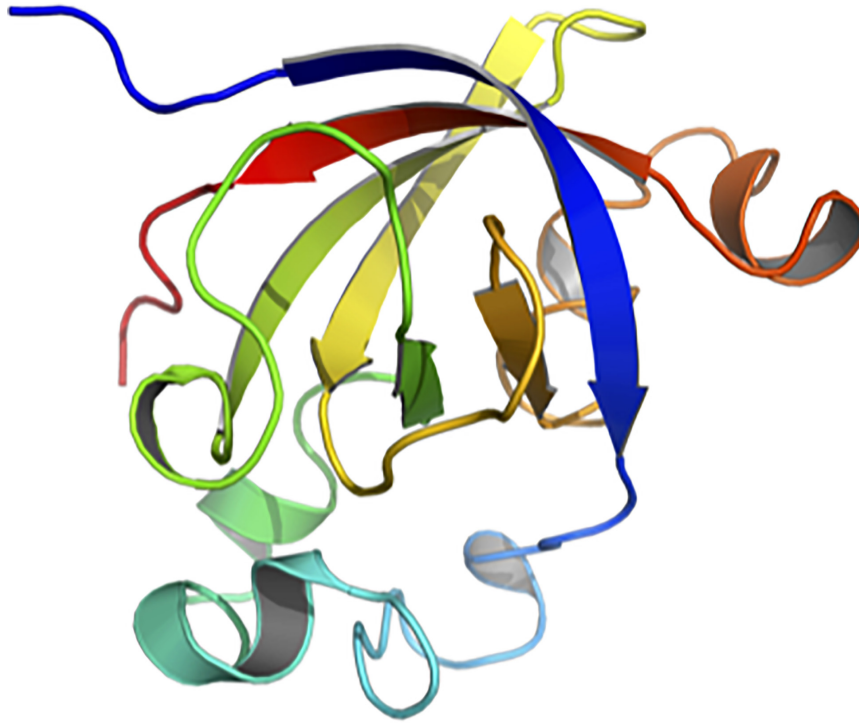
FIGURE 2 | Crystallographic structure of SUGARWIN2. The structure of SUGARWIN2 at 1.55 Å shown in cartoon-style using the "Chainbow" representation in which the N-terminus is dark blue and the C-terminus red. SUGARWIN2 has the classical BARWIN-like fold. The variable loop between the first and second β-strands (blue and green respectively) is two residues shorter than in previously reported structures.
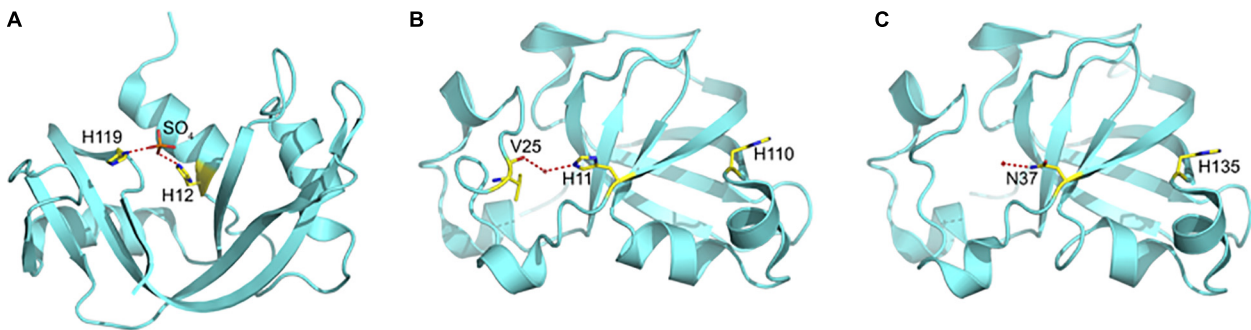

FIGURE 3 | The Catalytic Site in **(A)** Ribonusease A, **(B)** CARWIN, and **(C)** SUGARWIN2. In RNaseA, the catalytic histidines are orientated toward one another and bound to a sulfate ion which mimics the substrate where hydrolytic attack occurs. In CARWIN, although present, the histidines are orientated away from one another and the side chain of His11 is connected to the insertion within the variable loop by water mediated hydrogen bonds. In order to cleave RNA, it is expected that RNase active enzymes (including SUGARWIN1) must undergo a large conformational change to bring the two histidines into the correct orientation for catalysis. The presence of Asn37 in SUGARWIN2, in place of the histidine, makes this protein RNase inactive.

the asparagine side chain appears not to be fixed in a unique conformation but rather, adopts different rotamers in the two monomers of the asymmetric unit. However, in both of them it is structurally disconnected from the variable loop which lies further away due to the lack of the two-residue insertion.

SUGARWIN1 and SUGARWIN2 sequence analysis reveals several substitutions affecting charged amino acid residues. The majority of these alterations correspond to the substitution of basic residues in SUGARWIN1 by neutral residues in

SUGARWIN2 (**Figure 4**). These changes have a dramatic consequence for the pI of the two proteins: 8.78 for SUGARWIN1 and 4.82 for SUGARWIN2. By generating a homology model for SUGARWIN1 based on the crystal structure of SUGARWIN2 it was possible to map these residues onto the two molecular surfaces. **Figure 5** shows that they give rise to significant differences to the surface electrostatic potential. Such differences would be expected to have consequences for the binding partners of the two proteins and the molecular recognition

processes which govern their interaction. For example, the binding of RNA would be expected to be facilitated in the case of SUGARWIN1 compared with SUGARWIN2 based on electrostatic considerations. These changes, together with the absence of the catalytic histidine (His11), help to explain the lack of RNase activity of the latter.

Molecular modeling was also used to build 3D structures for these recently described SUGARWINs and the electrostatic surface analysis shows a mainly negative surface predicting a weak binding to RNA (**Figure 4**). Molecular modeling electrostatic surface analysis were also used to build 3D structures for the newly-identified SUGARWIN3 and SUGARWIN4. Our data shows a mainly negative surface predicting a weak binding to RNA (**Figure 4**). Future research is necessary to functionally characterize these SUGARWINs, however, we suggest that the lack of a complete RNase active site and a less positively charged molecular surface may go hand-in-hand.

Molecular Dynamics

The observation that SUGARWIN2 has a two residues deletion in the variable loop, which is connected by hydrogen bonds to the active site His11 in active RNases, stimulated us to investigate the dynamics of the molecule as a whole. This was further motivated by the observation, that reorientation of the catalytic residues would be expected to occur during catalysis. In order to analyze the consequences of this deletion on structural mobility, a molecular dynamics approach was employed, starting from the crystal structure of SUGARWIN2. Molecular modeling of the remaining 3 SUGARWINs (SUGARWIN1, SUGARWIN3 and SUGARWIN4) was also performed.

The molecular dynamics of the crystal structure of SUGARWIN2 together with the derived models for SUGARWINs 1, 3 and 4 shows a very similar pattern of intrinsic movements in all cases. The most notable features of this analysis are the large peaks associated with the variable loop in the model for SUGARWIN1 (residues 42–61) which are absent from the remaining structures. The RMSF data from the molecular dynamics simulations (**Figure 6**) highlight the mobility of the variable loop, as shows a major peak for SUGARWIN1 along residues 44–58. The loop begins at residue H11 [42—SUGARWIN1] which is believed to be involved in RNase activity and which, in the structure of CARWIN (H11), is coupled to an insertion within the loop itself by water mediated hydrogen bonds. SUGARWIN2 was the least flexible structure found in this analysis, with a minor peak between residues 60–65. SUGARWIN3 showed a major peak between residues 132 and 140, a region which precedes the second catalytic residue assumed to be essential for RNA hydrolysis (His113). SUGARWIN3 is the only structure to show this mobility and is also unique in having an asparagine at this position. Together with the behavior observed for SUGARWIN1 this suggests that structural mobility close to the catalytically implicated residues, H11 and H113, may be correlated with the type of amino acid which occupies these positions. SUGARWIN4 shows a single major peak involving residues 62–68, whereas other less prominent peaks are observed in all structures.

PCA analysis for each simulation was performed in order to identify putative sequence movements, which were then grouped and graphically represented. This data allowed the observation of the most flexible regions of each structure which correspond to the most representative group for each simulation (PC1), and are shown in **Figure 7**. Regarding the active site of the chitosanase function, only SUGARWIN1 showed significant movement in this region due to the flexibility of the variable loop described above (**Figure 8**). The mobility of this region may therefore be related to the ability for SUGARWIN1 to act as both a chitosanase and a ribonuclease, whereby different catalytic activities would be related to different loop conformations. We suggest that it is the dynamic nature of SUGARWIN1 (particularly the variable loop), which allows functional diversity implying that different structural conformations are a necessary requirement for different catalytic activities (**Figure 8**).

DISCUSSION

Plant response to biotic stresses such as herbivore insects and fungal pathogens involves a wide set of complex interactions, which cause profound physiological, chemical and morphological adaptations (Silva-Filho and Vivanco, 2017). Previous studies using transcriptomic data showed only two genes encoding for PR-4 proteins in sugarcane: SUGARWIN1 and SUGARWIN2 (Medeiros et al., 2012; Franco et al., 2014, 2019). However, in this work, we expand this observation by two additional protein sequences with distinct structural characteristics, named SUGARWIN3 and SUGARWIN4.

SUGARWIN1 exhibits RNase, chitinase and chitosanase activity, whereas SUGARWIN2 has only chitosanase activity (Franco et al., 2019). SUGARWIN3 and SUGARWIN4 have yet to be functionally characterized. Therefore, in order to shed light on the known differences among this protein family, we determined the three-dimensional structure of SUGARWIN2, which was crystallized and diffracted to 1.55 Å. Although SUGARWIN2 does not have RNase activity, its crystal structure shows the classic BARWIN fold (a six stranded β-sheet surrounded by four α-helices and several loops). In addition, we were able to build homology models for the remaining SUGARWINs for structural comparisons.

Examination of the structures for the four variants revealed that SUGARWIN1 presented at least three unique properties. Firstly, it is the only SUGARWIN to possess the catalytic residue, His11, associated with RNase activity; secondly, it is the most flexible of the four structures and finally, it has the most positive electrostatic surface potential. We argue that this unique combination of features renders SUGARWIN1 an active ribonuclease.

Several studies have correlated the RNase activity of BARWIN with the presence of two highly conserved histidine residues: at positions 11 and 113 (in relation to the BARWIN mature protein) [CARWIN (H11 and H110), WHEATWIN (H11 and H113) and SUGARWINs as referred above] (Caporale et al., 2004; Bertini et al., 2009; Huet et al., 2013; Franco et al., 2019). In addition, Bertini et al. (2009) also evaluated the relationship
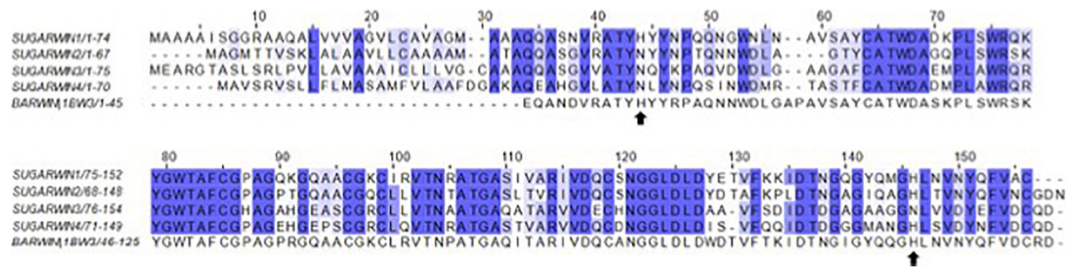
FIGURE 4 | Alignment of SUGARWINs and BARWIN. The alignment (ClustalW) of the full sequences, including the signal peptide, highlights a histidine at alignment position 44 (His11) in BARWIN and SUGARWIN1. SUGARWINs 2, 3 and 4 show the replacement of histidine by asparagine at alignment position 44.
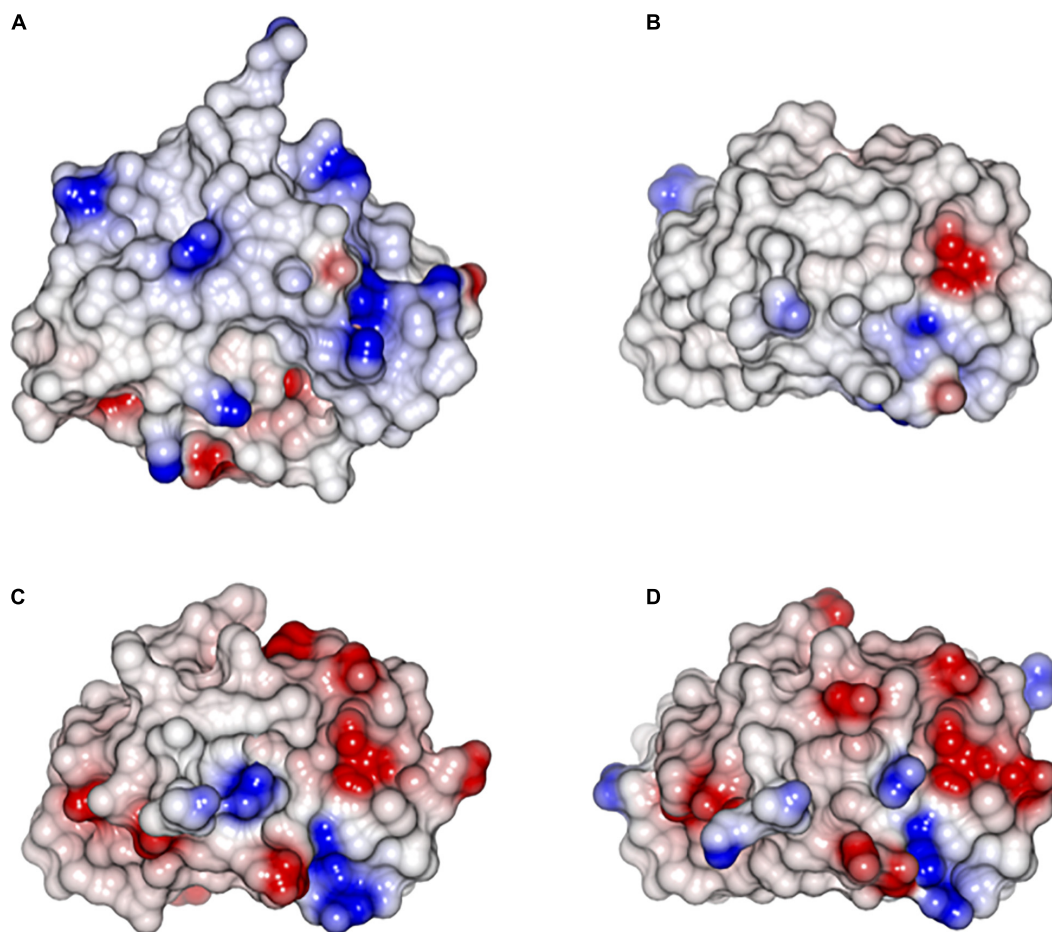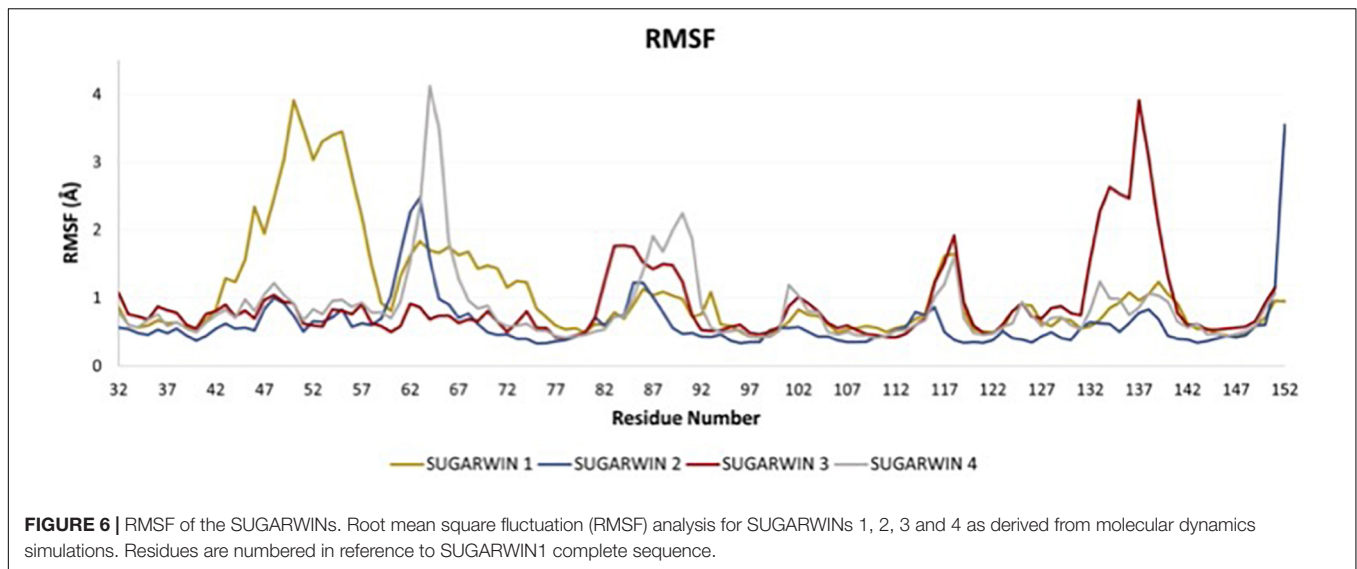

FIGURE 5 | Electrostatic surface of the SUGARWINs. **(A)** SUGARWIN1; **(B)** SUGARWIN2; **(C)** SUGARWIN3; **(D)** SUGARWIN4. SUGARWIN1 is more –positive (blue) than the remaining SUGARWINs. The electrostatic surface analysis shows a mainly electronegative surface in SUGARWIN 2, 3 and 4 suggestive of poor binding to RNA, if any.

between the ribonuclease activity of SUGARWIN counterparts and their antifungal properties. It has been demonstrated that two different mutations to His11 (H11G and H11L) of WHEATWIN partially inhibited RNase activity, pointing out the importance of this residue for catalysis. Furthermore, these authors also shown that the presence of the catalytic residues (His11 and His113, in WHEATWIN1) is also fundamental for the antifungal activity, at least in WHEATWIN1, thereby correlating the two phenomena. Huet et al. (2013) also reported that the two homologous histidine residues are important for RNase activity in CARWIN.

Although SUGARWINs 1, 2 and 4 all have His146, only SUGARWIN1 possesses both histidine residues. Franco et al. (2019) demonstrated the loss of SUGARWIN1's RNase activity by mutating His11 to asparagine (H42N—SUGARWIN1). This

FIGURE 6 | RMSF of the SUGARWINs. Root mean square fluctuation (RMSF) analysis for SUGARWINs 1, 2, 3 and 4 as derived from molecular dynamics simulations. Residues are numbered in reference to SUGARWIN1 complete sequence.
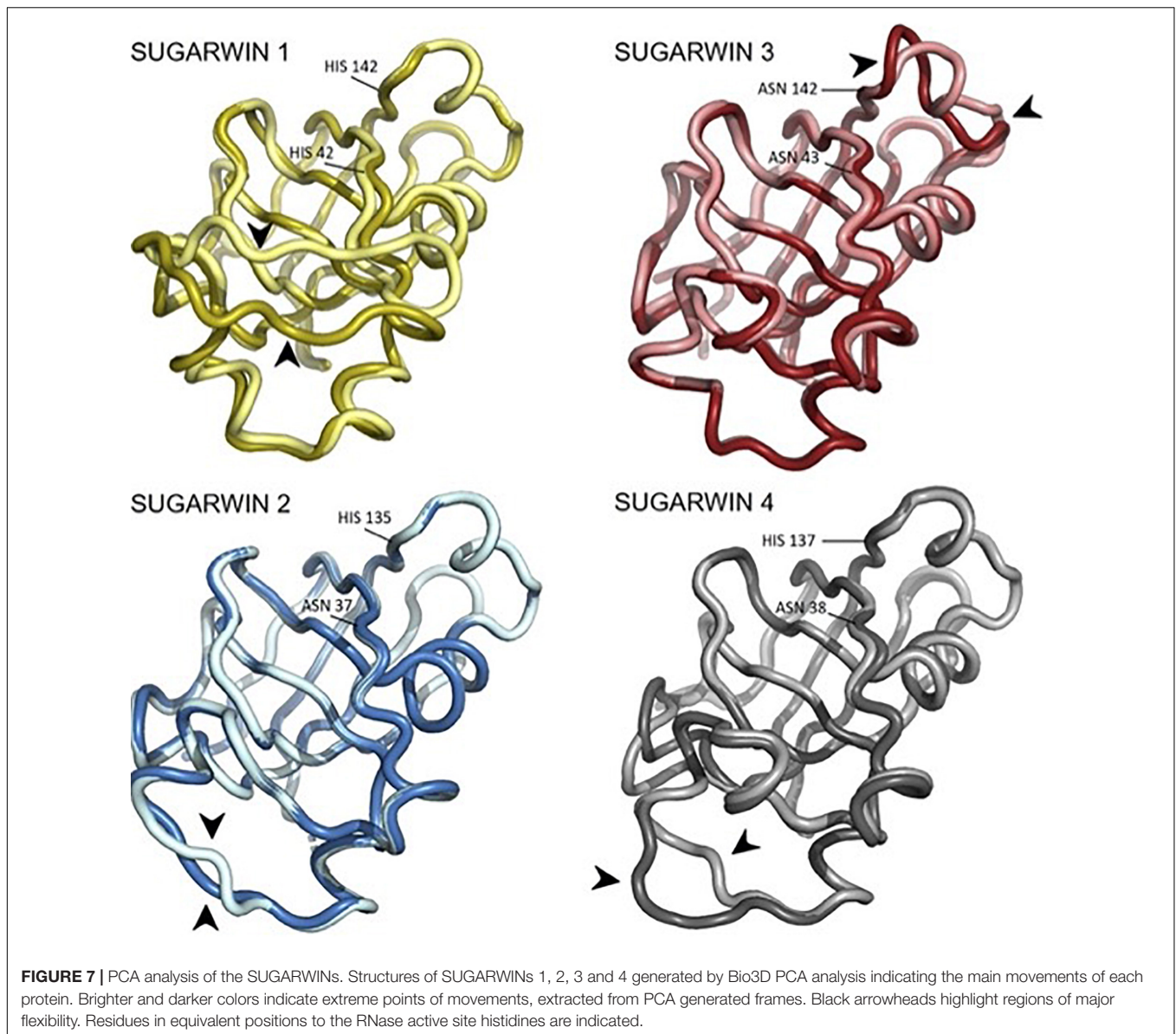
strongly suggests that RNA hydrolysis by SUGARWIN1 occurs according to the classic acid-base mechanism involving both histidines, as seen in RNase A, RNase T1, BARWIN, CARWIN and WHEATWINs (Silva-Filho and Vivanco, 2017).

Although SUGARWIN1 has the catalytic machinery necessary for the hydrolysis of RNA, the SUGARWIN2 structure, similar to CARWIN's, shows that the His residues at positions 11 and 113 (11 and 111—CARWIN; 37 and 135—SUGARWIN2) point to opposite directions. In the case of SUGARWIN1 the two catalytic histidines would be expected to be oriented similarly, as observed in our homology-based model. However, this arrangement is incompatible with the acid-base mechanism, which suggests that SUGARWIN1 must have sufficient intrinsic flexibility in order to reposition the catalytic diad in such a way as to be competent for catalysis. This hypothesis is compatible with the separation observed between the Cα atoms of Asn11 and His113 in SUGARWIN2 (10.6 Å), which is similar to that observed in bovine RNase A (8.9 Å), suggesting a putative rearrangement of the protein structure. For this reason, we investigated the dynamics of all four SUGARWIN structures with particular interest in SUGARWIN1.

SUGARWIN1 displayed significant flexibility throughout the entire region of the variable loop (residues 44–56), but particularly in its N-terminal portion (residues 44–51). This loop starts at the catalytic His11, which, as mentioned above, is only present in SUGARWIN1. Furthermore, this loop includes a two-residue insertion in SUGARWIN1 and this insertion is coupled to His11 *via* water mediated hydrogen bonds (**Figure 3**). It seems reasonable to assume that the altered dynamic behavior of the variable loop in SUGARWIN1 is due to these notable structural differences. A direct association between loop flexibility with RNase function has already been suggested, whereby the loop may be important for the binding and correct positioning of RNA in the active site as well as for RNA release after catalysis (Doucet et al., 2009; Menezes et al., 2014; Li and Hammes-Schiffer, 2019).

In contrast, SUGARWINs 2 and 4 showed significant structural mobility only within the C-terminal portion of the variable loop (residues 60–68, **Figure 6**), whilst SUGARWIN3 showed a markedly different profile altogether with a peak toward the C-terminus (residues 132–141), which precedes the latter catalytic histidine. It is worth emphasizing that SUGARWIN3, which is unique in presenting increased mobility in this region, is the only SUGARWIN that has an asparagine residue in place of histidine at this position (H142N—SUGARWIN3). Overall, it would therefore appear that differences in chain mobility would be related to the nature of the amino acids occupying the catalytic positions 11 and 113 as well as to the variable loop insertion in the case of SUGARWIN1. These results are in agreement with previous data (Ludvigsen and Poulsen, 1992; Bertini et al., 2009, 2012; Huet et al., 2013; Franco et al., 2014, 2019).

All known PR-4 ribonucleases, such as BARWIN, CARWIN and WHEATWIN, possess the two-residue insertion in the variable loop (Doucet et al., 2009; Menezes et al., 2014; Li and Hammes-Schiffer, 2019). Movement of this loop may be critical to give His11 the necessary structural liberty to move from the position seen in our model for SUGARWIN1 (and CARWIN) in order to form the RNase active site. This seems particularly evident on examining the CARWIN structure in which one of the two inserted residues of the loop holds the histidine sidechain to its mainchain carbonyl *via* a water-mediated hydrogen bond. The additional flexibility associated with the variable loop in SUGARWIN1 may be sufficient to decouple His11 from the loop insertion, thereby allowing it to adopt an alternative conformation—one compatible with RNase hydrolysis. However, we should point out that during our simulations we did not observe His11 pointing toward His113 as might be expected. This may suggest that the generation of a mature RNase active site may depend not only on the flexibility of the variable loop but also on association with the substrate itself, RNA, through an induced-fit like mechanism.

FIGURE 7 | PCA analysis of the SUGARWINs. Structures of SUGARWINs 1, 2, 3 and 4 generated by Bio3D PCA analysis indicating the main movements of each protein. Brighter and darker colors indicate extreme points of movements, extracted from PCA generated frames. Black arrowheads highlight regions of major flexibility. Residues in equivalent positions to the RNase active site histidines are indicated.
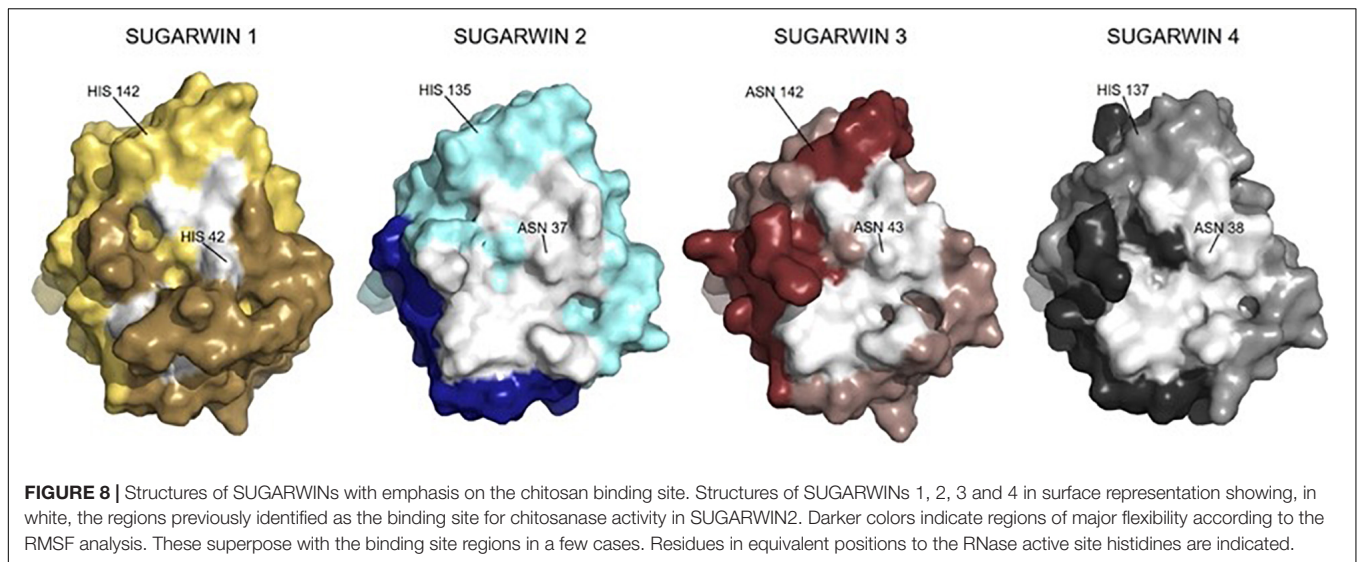
Evidence for the reorientation of His11 comes from the NMR structure of BARWIN, where it is partially rotated in the direction of His113, its partner in the RNase active site. This is facilitated by the variable loop adopting a different conformation in BARWIN, such that His11 is not coupled to the insertion within the variable loop. This is a direct evidence that changes to the conformation of the variable loop could be related to liberating His11 in order to allow the formation of a mature RNase active site. Amongst the four different variants only SUGARWIN1 has the histidine at position 11 and the necessary structural plasticity.

For the ribonuclease activity of SUGARWIN1 to occur it is not sufficient to have the catalytic machinery and the expected flexibility allowing the formation of the active site with the histidines residues properly oriented. In addition to these two features, it is also necessary for the enzyme to attract and bind the highly negatively charged substrate, RNA. Examination of

the electrostatic potential of the four SUGARWINs revealed that SUGARWIN1 is, by far, the most basic protein. This is probably due to a series of amino acid substitutions leading to an accumulation of lysines and arginines residues on the molecular surface (**Figure 3**) and a significant increase in the pI in comparison to the remaining SUGARWINs. This alteration to the electrostatic properties of the SUGARWIN1 surface would be expected to back the formation of the enzyme-substrate complex thereby permitting RNA hydrolysis to occur.

Different from the RNase activity, both SUGARWIN1 and SUGARWIN2 are known to be active chitosanases. However, the major movement to the variable loop in SUGARWIN1 leads to a significant occlusion of the known chitosan binding site (**Figure 8**). Therefore, the loop movement which we consider to be critical for the RNase active site, simultaneously covers the chitosan binding site. This suggests that SUGARWIN1 (and

FIGURE 8 | Structures of SUGARWINs with emphasis on the chitosan binding site. Structures of SUGARWINs 1, 2, 3 and 4 in surface representation showing, in white, the regions previously identified as the binding site for chitosanase activity in SUGARWIN2. Darker colors indicate regions of major flexibility according to the RMSF analysis. These superpose with the binding site regions in a few cases. Residues in equivalent positions to the RNase active site histidines are indicated.

by analogy BARWIN, CARWIN and WHEATWIN) would be unable to simultaneously perform both catalytic functions. If the variable loop is positioned to expose the chitosan-binding site, then the RNase active site is not formed, as implied by the structure for SUGARWIN2. On the other hand, it seems likely that when the latter is correctly formed, then the chitosan-binging site will be occluded by the loop. Therefore, we propose that SUGARWIN1 might be considered as an enzyme with two different catalytic functions which are mutually exclusive.

No function has been described yet for the newly-described SUGARWINs (SUGARWIN3 and 4). It seems clear that no nuclease function might be associated with these proteins, since both lack the catalytic histidine at position 11. Additionally, SUGARWIN3 also has an asparagine residue replacing the histidine at position 113. Nevertheless, according to the structural analyses, it might be expected that both proteins display chitosanase activities. This is consistent with the fact that it is only in SUGARWIN1 that we observe the variable loop crosses over the chitosan binding site. For the other isoforms, the binding site for the chitosan substrate remains exposed during the simulations. We have previously shown that for SUGARWIN1 both the RNase and chitinase activities do not impair its chitosanase function. As mentioned above this is probably associated with the structural flexibility in the region of the variable loop which is most notable in SUGARWIN1, the only variant with a known nuclease function (SUGARWIN1).

Our phylogenetic analysis showed that both histidines related to RNase activity and a variable loop of the same size to that found in the variable region of SUGARWIN1 are present in the majority of the Embryophyta PR-4 protein sequences and are widespread in the phylogenetic tree. In contrast, proteins with an asparagine residue at position 11 were found only in a few genomes and do not form a separate clade. These results show that both histidine and the flexible loop observed in SUGARWIN1 may have been present in the PR-4 protein from the Embryophyta ancestral species and that independent replacements by asparagine may have happened in the evolutionary history of this group. One of

these replacements may have occurred in the common ancestor of Poales originating the sequence that gave rise to SUGARWIN2, SUGARWIN3 and SUGARWIN4 (**Figure 1**).

CONCLUSION

Thus far only two PR-4 proteins have had their tridimensional structures solved: BARWIN from Barley [PDB ID: 1BW3 (Ludvigsen and Poulsen, 1992)] and CARWIN from papaya [PDB ID: 4JP6 (Huet et al., 2013)], both of which contain H11 and H113 (H11 and H110 in CARWIN) and with chitinase and RNase activity. Here, we described the first tridimensional structure of a PR-4 protein with no RNase activity and with the H44N replacement: SUGARWIN2. Comparison of the structures, dynamics and surface properties of the two proteins allows us to propose that (1) the more positive surface potential of SUGARWIN1 aids in binding the substrate RNA, (2) flexibility of the variable loop is essential for the formation of the RNase binding site which otherwise would have the catalytic histidines pointing to opposite directions as observed in the crystal structures, and (3) the flexibility of the loop also occludes the chitosan binding site suggesting that the two catalytic activities are mutually exclusive. Unexpectedly, despite its lack of RNase activity, SUGARWIN2 seems to be the more relevant for sugarcane defense against pathogenic fungi, since its expression is strongly upregulated after 10 days of *C. falcatum* treatment, whereas SUGARWIN1 showed a moderate increase in gene expression under the same conditions. This may imply that a more specialized enzyme is required for this purpose. We also describe here two new SUGARWIN proteins, SUGARWIN3 and SUGARWIN4, which are predicted, based on the above analyses, to lack RNase activity. Further studies are being conducted to better understand these new proteins. The exact details of the catalytic capabilities of each of these enzymes appears to be a complex function of several factors including the residues which comprise the active site, the intrinsic flexibility of the

structures themselves and the surface electrostatics necessary for substrate recognition.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

MS-F: design of the experiments, lab infrastructure, providing research funds, discussions, and writing. All authors contributed to the article and approved the submitted version.

REFERENCES

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.734248/full#supplementary-material

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Echols, N., Headd, J. J., et al. (2011). The Phenix software for automated determination of macromolecular structures. *Methods* 55, 94–106. doi: 10.1016/j.ymeth.2011.07.005

Agrawal, G. K., Jwa, N. S., Han, K. S., Agrawal, V. P., and Rakwal, R. (2003). Isolation of a novel rice PR4 type gene whose mRNA expression is modulated by blast pathogen attack and signaling components. *Plant Physiol. Biochem.* 41, 81–90. doi: 10.1016/s0981-9428(02)00012-8

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

Bai, S., Dong, C., Li, B., and Dai, H. (2013). A PR-4 gene identified from *Malus domestica* is involved in the defense responses against *Botryosphaeria dothidea*. *Plant Physiol. Biochem.* 62, 23–32. doi: 10.1016/j.plaphy.2012.10.016

Banno, S., Ochiai, N., Noguchi, R., Kimura, M., Yamaguchi, I., Kanzaki, S., et al. (2005). A catalytic subunit of cyclic AMP-dependent protein kinase, PKAC-1, regulates asexual differentiation in *Neurospora crassa*. *Genes Gen. Syst.* 80, 25–34. doi: 10.1266/ggs.80.25

Bertini, L., Caporale, C., Testa, M., Proietti, S., and Caruso, C. (2009). Structural basis of the antifungal activity of wheat PR4 proteins. *FEBS Lett.* 583, 2865–2871. doi: 10.1016/j.febslet.2009.07.045

Bertini, L., Proietti, S., Aleandri, M. P., Mondello, F., Sandini, S., Caporale, C., et al. (2012). Modular structure of HEL protein from *Arabidopsis* reveals new potential functions for PR-4 proteins. *Biol. Chem.* 393, 1533–1546. doi: 10.1515/hsz-2012-0225

Bravo, J. M., Campo, S., Murillo, I., Coca, M., and Segundo, B. S. (2003). Fungusand wound-induced accumulation of mRNA containing a class II chitinase of the pathogenesis-related protein 4 (PR-4) family of maize. *Plant Mol. Biol.* 52, 745–759.

Broekaert, I., Lee, H. I., Kush, A., Chua, N. H., and Raikhel, N. (1990). Woundinduced accumulation of mRNA containing a hevein sequence in laticifers of rubber tree (*Hevea brasiliensis*). *Proc. Natl. Acad. Sci. U.S.A.* 87, 7633–7637. doi: 10.1073/pnas.87.19.7633

Capella-Gutierrez, S., Silla-Martinez, J. M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Caporale, C., Di Berardino, I., Leonardi, L., Bertini, L., Cascone, A., Buonocore, V., et al. (2004). Wheat pathogenesis-related proteins of class 4 have ribonuclease activity. *FEBS Lett.* 575, 71–76. doi: 10.1016/j.febslet.2004.07.091

Caruso, C., Bertini, L., Tucci, M., Caporale, C., Leonardi, L., Saccardo, F., et al. (1999). Isolation and characterisation of wheat cDNA clones encoding PR4 proteins. *DNA Seq.* 10, 301–307. doi: 10.3109/10425179909033956

Caruso, C., Nobile, M., Leonardi, L., Bertini, L., Buonocore, V., and Caporale, C. (2001). Isolation and amino acid sequence of two new PR-4 proteins from wheat. *J. Protein Chem.* 20, 327–335.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., et al. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 66(Pt 1), 12–21. doi: 10.1107/s0907444909042073

Christopher, M. E., Miranda, M., Major, I. T., and Constabel, C. P. (2004). Gene expression profiling of systemically wound-induced defenses in hybrid poplar. *Planta* 219, 936–947. doi: 10.1007/s00425-004-1297-3

Doucet, N., Watt, E. D., and Loria, J. P. (2009). The flexibility of a distant loop modulates active site motion and product release in ribonuclease A. *Biochemistry* 48, 7160–7168. doi: 10.1021/bi900830g

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60(Pt 12 Pt 1), 2126–2132. doi: 10.1107/s0907444904019158

Falco, M. C., Marbach, P. A. S., Pompermayer, P., Lopes, F. C. C., and Silva-Filho, M. C. (2001). Mechanisms of sugarcane response to herbivory. *Genet. Mol. Biol.* 24, 113–122. doi: 10.1590/s1415-47572001000100016

Franco, F. P., Dias, R. O., Toyama, D., Henrique-Silva, F., Moura, D. S., and Silva-Filho, M. C. (2019). Structural and functional characterization of PR-4 SUGARWINs from sugarcaneand their role in plant defense. *Front. Plant Sci.* 9:1916. doi: 10.3389/fpls.2018.01916

Franco, F. P., Moura, D. S., Vivanco, J. M., and Silva-Filho, M. C. (2017). Plant-insect-pathogen interactions: a naturally complex menage a trois. *Curr. Opin. Microbiol.* 37, 54–60. doi: 10.1016/j.mib.2017.04.007

Franco, F. P., Santiago, A. C., Henrique-Silva, F., de Castro, P. A., Goldman, G. H., Moura, D. S., et al. (2014). The sugarcane defense protein SUGARWIN2 causes cell death in Colletotrichum falcatum but not in non-pathogenic fungiThe sugarcane defense protein SUGARWIN2 causes cell death in Colletotrichum falcatum but not in non-pathogenic fungi. *PLoS One* 9:e91159. doi: 10.1371/journal.pone.0091159

Franco, F. P., Túler, A. C., Gallan, D. Z., Gonçalves, F. G., Favaris, A. P., Peñaflor, M. F. G. V., et al. (2021). Fungal phytopathogen modulates plant and insect responses to promote its dissemination. *ISME J.* 15:10. doi: 10.1038/s41396-021-01010-z

Friedrich, L., Moyer, M., Ward, E., and Ryals, J. (1991). Pathogenesis-related protein-4 is structurally homologous to the carboxy-terminal domains of hevein, win-1 and win-2. *Mol. Gen. Genet.* 230, 113–119. doi: 10.1007/bf00290658

Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A., and Caves, L. S. D. (2006). Bio3D: an R package for the comparative analysis of protein structures. *Bioinformatics* 22, 2695–2696. doi: 10.1093/bioinformatics/btl461

Guevara-Morato, M., de Lacoba, M. G., Garcia-Luque, I., and Serra, M. T. (2010). Characterization of a pathogenesis-related protein 4 (PR-4) induced in Capsicum chinense L-3 plants with dual RNase and DNase activities. *J. Exp. Bot.* 61, 3259–3271. doi: 10.1093/jxb/erq148

Hejgaard, J., Jacobsen, S., Bjorn, S. E., and Kragh, K. M. (1992). Antifungal activity of chitin-binding pr-4 type proteins from barley-grain and stressed leaf. *FEBS Lett.* 307, 389–392. doi: 10.1016/0014-5793(92)80720-2

Huet, J., Mbosso, E. J. T., Soror, S., Meyer, F., Looze, Y., Wintjens, R., et al. (2013). High resolution structure of a papaya plant-defence barwin-like protein solved by in-house sulfur-SAD phasing. *Acta Crystallogr. Sec. D Biol. Crystallogr.* 69, 2017–2026. doi: 10.1107/s0907444913018015

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 3, 1236–1240. doi: 10.1093/bioinformatics/btu031

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kiba, A., Saitoh, H., Nishihara, M., Omiya, K., and Yamamura, S. (2003). C-terminal domain of a hevein-like protein from Wasabia japonica has potent antimicrobial activity. *Plant Cell Physiol.* 44, 296–303. doi: 10.1093/pcp/pcg035

Kim, N. H., and Hwang, B. K. (2015). Pepper pathogenesis-related protein 4c is a plasma membrane-localized cysteine protease inhibitor that is required for plant cell death and defense signaling. *Plant J.* 81, 81–94. doi: 10.1111/tpj.12709

Li, P., and Hammes-Schiffer, S. (2019). Substrate-to-product conversion facilitates active site loop opening in yeast enolase: a molecular dynamics study 9. *ACS Catal.* 9, 8985–8990. doi: 10.1021/acscatal.9b03249

Li, X. D., Xia, B., Jiang, Y. M., Wu, Q. S., Wang, C. Y., He, L. S., et al. (2010). A new pathogenesis-related protein, LrPR4, from *Lycoris radiata*, and its antifungal activity against *Magnaporthe grisea*. *Mol. Biol. Rep.* 3, 995–1001. doi: 10.1007/s11033-009-9783-0

Linthorst, H. J. M., Danhash, N., Brederode, F. T., Vankan, J. A. L., Dewit, P., and Bol, J. F. (1991). Tobacco and tomato pr proteins homologous to win and pro-hevein lack the hevein domain. *Mol. Plant Microbe Interact.* 4, 586–592. doi: 10.1094/mpmi-4-586

Ludvigsen, S., and Poulsen, F. M. (1992). Positive ϕ-angles in proteins by nuclear magnetic resonance spectroscopy. *J. Biomol. NMR* 2, 227–233. doi: 10.1007/bf01875318

Marti-Renom, M. A., Stuart, A., Fiser, A., Sánchez, R., Melo, F., and Sali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* 29, 291–325. doi: 10.1146/annurev.biophys.29.1.291

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Cryst.* 40, 658–674.

Medeiros, A. H., Franco, F. P., Matos, J. L., de Castro, P. A., Santos-Silva, L. K., Henrique-Silva, F., et al. (2012). Sugarwin: a sugarcane insect-induced gene with antipathogenic activity. *Mol. Plant Microbe Interact.* 25, 613–624. doi: 10.1094/mpmi-09-11-0254

Menezes, S. P., de Andrade Silva, E. M., Matos Lima, E., Oliveira de Sousa, A., Andrade, B. S., Lemos, L. S. L., et al. (2014). The pathogenesis-related protein PR-4b from *Theobroma cacao* presents RNase activity, Ca2+ and Mg2+ dependentDNase activity and antifungal action on *Moniliophthora perniciosa*. *BMC Plant Biol.* 14:161. doi: 10.1186/1471-2229-14-161

Neuhaus, J. M., Fritig, B., Linthorst, H. J. M., Meins, F., Mikkelsen, J. D., and Ryals, J. (1996). A revised nomenclature for chitinase genes. *Plant Mol. Biol. Rep.* 14, 102–104. doi: 10.1007/bf02684897

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. I. (2015). A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Olsson, M. H., Søndergaard, C. R., Rostkowski, M., and Jensen, J. H. (2011). Propka3: consistent treatment of internal and surface residues in empirical p k a predictions. *J. Chem. Theory Comput.* 7, 525–537. doi: 10.1021/ct100578z

Pieterse, C. M. J., and van Loon, L. C. (1999). Salicylic acid-independent plant defence pathways. *Trends Plant Sci.* 4, 52–58. doi: 10.1016/s1360-1385(98)01364-8

Potter, S., Uknes, S., Lawton, K., Winter, A. M., Chandler, D., Dimaio, J., et al. (1993). Regulation of a hevein-like gene in *Arabidopsis*. *Mol. Plant Microbe Interact.* 6, 680–685. doi: 10.1094/mpmi-6-680

Ryan, C. A. (1990). Protease Inhibitors in plants—genes for improving defenses against insects and pathogens. *Annu. Rev. Phytopathol.* 28, 425–449. doi: 10.1146/annurev.py.28.090190.002233

Schilmiller, A. L., and Howe, G. A. (2005). Systemic signaling in the wound response. *Curr. Opin. Plant Biol.* 8, 369–377. doi: 10.1016/j.pbi.2005.05.008

Silva-Filho, M. C., and Vivanco, J. M. (2017). Guest editorial: plants and their surrounding microorganisms: a dynamic world of interactions. *Curr. Opin. Microbiol.* 37, 5–6.

Souza, G. M., Van Sluys, M. A., Lembke, C. G., Lee, H., Margarido, G. R. A., Hotta, C. T., et al. (2019). Assembly of the 373k gene space of the polyploid sugarcane genome reveals reservoirs of functional diversity in the world's leading biomass crop. *Gigascience* 8:giz129. doi: 10.1093/gigascience/giz129

Souza, T. P., Dias, R. O., and Silva-Filho, M. C. (2017). Defense-related proteins involved in sugarcane responses to biotic stress. *Genet. Mol. Biol.* 40, 360–372. doi: 10.1590/1678-4685-gmb-2016-0057

Stein, N. (2008). CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Cryst.* 41, 641–643. doi: 10.1107/s0021889808006985

Svensson, B., Svendsen, I., Hojrup, P., Roepstorff, P., Ludvigsen, S., and Poulsen, F. M. (1992). Primary structure of BARWIN - a barley seed protein closely related to the c-terminal domain of proteins encoded by wound-induced plant genes. *Biochemistry* 31, 8767–8770. doi: 10.1021/bi00152a012

Van Loon, L. C., and Van Strien, E. A. (1999). The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. '*Physiol. Mol Plant Pathol.* 55, 85–97. doi: 10.1006/pmpp.1999.0213

Xu, Y., Chang, P. F. L., Liu, D., Narasimhan, M. L., Raghothama, K. G., Hasegawa, P. M., et al. (1994). Plant defense genes are synergistically induced by ethylene and methyl jasmonate. *Plant Cell* 6, 1077–1085. doi: 10.2307/3869886

Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., et al. (2018). Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat. Genet.* 50, 1565–1573.

Zhu, T., Song, F., and Zheng, Z. (2006). Molecular characterization of the rice pathogenesis-related protein, OsPR-4b, and its antifungal activity against *Rhizoctonia solani*. *J. Phytopathol.* 154, 378–384. doi: 10.1111/j.1439-0434.2006.01110.x

Check for updates

Conservation and Divergence of the *CONSTANS-Like (COL)* Genes Related to Flowering and Circadian Rhythm in *Brassica napus*

Yuxi Chen[1,2], Rijin Zhou[2], Qiong Hu[2], Wenliang Wei[1]* and Jia Liu[2]*

[1] College of Agriculture, Yangtze University, Jingzhou, China, [2] Oil Crops Research Institute of Chinese Academy of Agricultural Sciences/Key Laboratory for Biological Sciences and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, Wuhan, China

The *CONSTANS-LIKE (COL)* genes are important signaling component in the photoperiod pathway and flowering regulation pathway. However, people still know little about their role in *Brassica napus*. To achieve a better understanding of the members of the *BnaCOL* gene family, reveal their evolutionary relationship and related functions involved in photoperiod regulation, we systematically analyzed the *BnaCOL* family members in *B. napus* genome. A total of 33 *BnaCOL* genes distributed unevenly on 16 chromosomes were identified in *B. napus* and could be classified into three subfamilies. The same subfamilies have relatively conservative gene structures, three-dimensional protein structures and promoter motifs such as light-responsive *cis*-elements. The collinearity analysis detected 37 pairs of repetitive genes in *B. napus* genome. A 67.7% of the *BnaCOL* genes were lost after *B. napus* genome polyploidization. In addition, the *BnaCOL* genes showed different tissue-specific expression patterns. A 81.8% of the *BnaCOL* genes were mainly expressed in leaves, indicating that they may play a conservative role in leaves. Subsequently, we tested the circadian expression profiles of nine homologous genes that regulate flowering in *Arabidopsis*. Most *BnaCOL* genes exhibit several types of circadian rhythms, indicating that these *BnaCOL* genes are involved in the photoperiod pathway. As such, our research has laid the foundation for understanding the exact role of the *BnaCOL* family in the growth and development of rapeseed, especially in flowering.

Keywords: **Brassica napus**, **CONSTANS-LIKE (COL)** genes, genome-wide analysis, expression pattern, photoperiod

INTRODUCTION

Flowering is an important link in the process of plant reproduction (Fitter and Fitter, 2002). In *Arabidopsis thaliana,* the photoperiod pathway, vernalization pathway, autonomous pathway and gibberellin pathway constitute a complex genetic network that regulates flowering time (Roux et al., 2006). In *Arabidopsis*, *CONSTANS-like (CO/COL)* and *FLOWERING LOCUS T (FT)*

are two important network regulation centers in the photoperiod induction pathway. *CO/COL* activates the transcription of FT to move the FT protein from the leaf phloem to the shoot apex meristem, thereby promoting plant flowering (Shim et al., 2017). In rice, the *Arabidopsis CO/COL* homologous gene *HEADING DATE 1* (*Hd1*) appears to be a bifunctional regulator. It induces FT homologous *HEADING DATE 3a* (*Hd3a*) gene expression to promote flowering under SD conditions while under LD conditions *Hd1* functions as an inhibitor of *Hd3a* transcription and flowering (Hayama et al., 2003).

CO/COL is an important network center of the photoperiod flowering pathway, integrating together various environmental and internal signals (Shim et al., 2017). Structurally, *CO/COL* genes contain two conserved domains: a C-terminal CCT domain (also termed CO, CO-like, TOC1) and an N-terminal zinc finger B-box domain (Robson et al., 2001). The B-box domains are found in many kinds of animal proteins, including some transcription factors, ribonucleoprotein and proto-oncogene products (Reddy et al., 1992; Borden, 1998), and it acts as a protein-protein interaction domain in several transcription factors in animals (Borden, 1998). The CCT domain has the function of nuclear localization similar to the yeast HEME ACTIVATOR PROTEIN2 (HAP2) protein and participates in DNA binding (Wenkel et al., 2006). In *Arabidopsis*, the 17 *AtCOL* genes can be classified into three subgroups according to the difference in structural domains (Robson et al., 2001): (i) *CO, COL1-COL5* form the first subgroup, and they all have two B-box domains and one CCT domain. (ii) The second subgroup consists of *COL9-COL15* members. Compared with members in other groups, they have a zinc finger domain in addition to a B-box domain and a CCT domain. (iii) The third subgroup includes *COL6-COL8* and *COL16* with one B-box and one CCT domain (Robson et al., 2001; Griffiths et al., 2003). However, there are exceptions in structural domains, such as *OsH* and *OsI* in rice and *HvCO9* in barley, which contain an intron and a CCT domain, but lack the B-box structure (Robson et al., 2001; Griffiths et al., 2003).

The *COL* gene family have been widely studied in angiosperms, such as monocotyledonous plants (rice, barley, maize, etc.) (Griffiths et al., 2003; Song et al., 2018) and dicotyledonous plants (*Arabidopsis,* soybean, cotton, tomato, etc.) (Robson et al., 2001; Wu et al., 2014; Cai et al., 2017; Yang et al., 2020). The *COL* genes function in all developmental stages of plants (**Supplementary Table 1**). In *A. thaliana, AtCOL1-AtCOL2* have less effect on flowering time, but overexpression of *AtCOL1* can shorten the cycle of circadian rhythm, and have a certain impact on the light input pathway (Ledger et al., 2001). Both *AtCOL3* and *AtCOL4* play an inhibitory role in flowering in both LD and SD (Datta et al., 2006; Steinbach, 2019). Besides, *AtCOL3* also promotes red light signal transmission, lateral root growth, bud branching and anthocyanin accumulation (Datta et al., 2006), while mutation of *AtCOL4* shows increased tolerance to ABA and salt stress (Min et al., 2015). Recently, *AtCOL3* and *AtCOL13* were found to be co-regulator of hypocotyl elongation under red light (Liu B. et al., 2021). Overexpression of *AtCOL5* will bloom early under SD conditions (Hassidim et al., 2009). *AtCOL7* not only affects the branching and shade response of

A. thaliana (Wang et al., 2013), it is also a key factor linking light perception to auxin homeostasis (Zhang et al., 2014). And *AtCOL8* and *AtCOL9* transgenic plants flower late under LD conditions (Cheng and Wang, 2005; Takase et al., 2012). In rice, *OsCOL3, OsCOL4, OsCOL9, OsCOL10, OsCOL13, OsCOL15,* and *OsCOL16* act as flowering inhibitor to delay flowering time (Kim et al., 2008; Lee et al., 2010; Liu H. et al., 2016; Sheng et al., 2016; Tan et al., 2016, 2017; Wu et al., 2017, 2018), except *Ghd2*, which regulates leaf senescence and drought resistance (Liu J. et al., 2016). Furthermore, *CO/COL* genes are also found to regulate flowering time in potato (González-Schain and Suárez-López, 2008), sugar beets (Chia et al., 2008; Dally et al., 2018), soybean (Wu et al., 2014), sorghum (Yang et al., 2014), and bamboo (Xiao et al., 2018).

Brassica napus is an important and worldwide cultivated oil crop with strong adaptability, wide use and high economic value. Due to the great different cultivation in latitude, longitude and climate, different ecological types of *B. napus* varieties are needed. Hence, it is also a good plant resource to research flowering pattern and photoperiod rhythm like spring ecotype, winter ecotype and semi-winter ecotype. And the *CO/COL* genes are important transcription element in photoperiod and flowering regulation pathway. At present, the functions of *CO/COL* gene family members have been comprehensively studied in the model plant *Arabidopsis thaliana*, but little is known about *CO/COL* genes preservation and functional differentiation in *B. napus* after polyploidy events. In this study, we have identified 33 *BnaCOL* gene family members and performed bioinformatics analysis on their physical and chemical properties, evolutionary relationships, chromosome location, gene structure, three-dimensional protein structures, *cis*-acting elements of the promoter, GO annotation enrichment analysis and gene duplication. We also studied the expression patterns of *BnaCOL* gene family in different tissues and their response to SD or LD light treatment. This study would provide important clues for the functional study of the *COL* gene family in the Cruciferae plants, and lay a foundation for further exploration of its functional and molecular mechanisms.

MATERIALS AND METHODS

Identification of *CO-like* Transcription Factor Family in Rapeseed

The genome sequences, protein sequences and gene annotation files of rapeseed were downloaded from the website (BnPIR)[1] (Song et al., 2020). The Markov model of the two domains of CO-like CCT (PF06203) and zinc finger B-box (PF00643) was downloaded from Pfam database[2]. Using these two Markov models to preliminarily screen the protein sequences of rapeseed on the HMMER software, and the cut-off *E*-value were set to 1e-4, respectively. Subsequently, all candidate proteins were submitted

---

[1]http://cbi.hzau.edu.cn/bnapus/

[2]http://pfam.xfam.org/

to three online websites, i.e., SMART[3], NCBI CDD[4] and PFAM (see text footnote 2) to screen out candidate COL proteins with both CCT and B-box conserved domains. The identified *COL* candidate genes were submitted to the ExPASy website[5] for prediction analysis of protein molecular weight (MW) and isoelectric point (pI). And subcellular localization is predicted by WoLF PSORT[6].

Chromosome Location and Phylogenetic Analysis

The chromosome location data of *BnaCOLs* comes from the BnPIR website (see text footnote 1). And then the MapChart software was used to analyze the distribution of the identified *BnaCOLs* on rapeseed chromosomes. The results were refined with Adobe Illustrator software.

According to the reported literature (Hu et al., 2018), the protein sequences of *COL* family members of *Arabidopsis*, *B. oleracea*, *B. rapa*, *Capsella rubella*, *Oryza sativa*, *Raphanus sativus*, and *Zea mays* were downloaded and Clustal W was used to analyze the COL protein sequences of these plants. In addition, the sequence alignment results were submitted to MEGA 7.0 software, and the neighbor joining method (NJ) was used to construct the evolutionary tree (Saitou and Nei, 1987; Kumar et al., 2016).

Gene Structure and Protein Conservative Domain Analysis

The exon and intron structures of *COL* genes in rapeseed were analyzed by Gene Structure Display Server 2.0[7] (Hu et al., 2015). The BnaCOL protein sequences were submitted to the MEME software to analyze the conserved domain of genes. Setting the maximum number of motifs to 10, the maximum number of motif amino acids to 20 and the minimum width to 6 and other settings to default. Finally, TBtools software was used to visualize the conserved motifs of BnaCOL proteins.

Multi Sequence Alignment and Three-Dimensional Structure Prediction of Protein

We submitted 33 protein sequences of BnaCOL to DNAMAN7.0 software for multiple sequence comparison. Subsequently, we used the online website Phyre2[8] to predict the three-dimensional structure of the protein.

Collinearity Analysis Within *Brassica napus* and Among Different Species

The analysis of intra-species collinearity of *BnaCOL* genes in *B. napus* was performed with McScanX software and the relationship was plotted with Circos software. In addition,

the collinearity analysis was plotted with Python version of McScanX software.

Cis-Acting Element and Functional Annotation Analysis

The 1,500 bp upstream sequences of *BnaCOL* genes were obtained from *B. napus* Whole Genome Information Resource Website (see text footnote 1). The online website Plant CARE[9] was used to extract homeopathic a components, and then using the online website DSGS to visualize.

To shed light on the function of the *BnaCOL* genes, we used eggNOG database[10] for the gene ontology (GO) annotation analysis. Subsequently, the GO annotation data was processed in TBtools.

Tissue-Specific Expression Pattern of *BnaCOL* Genes

At the online website BnTIR: *Brassica napus* transcriptome information resource[11] (Liu D. et al., 2021), we downloaded RNA-seq data of different tissues including roots, cotyledons, leafs, sepals, petals, pollen, buds, siliques, and seeds. The data were submitted to the online tool[12] to draw the expression heat map.

Plant Materials and Treatment Methods

The seeds of Zhongshuang 11 were grown in a growth chamber with a temperature of 25°C/18°C, light for 16 h/darkness for 8 h and humidity of 80%. When the seedlings were at the five leaf stage, two different photoperiod treatments were applied: long daylight (LD, 16 h light/8 h dark) and short daylight (SD, 8 h light/16 h dark). We collected the third leaf of these seedlings at 0, 4, 8, 12, 16, 20, and 24 h after photoperiod treatment. Besides, we set up three biological replicates with samples collected. The collected leaves were immediately frozen in liquid nitrogen and then stored in −80°C refrigerator.

RNA Extraction and RT-PCR Analysis

Total RNA was extracted from leaves treated with different photoperiodic treatments using polysaccharide polyphenol total RNA extraction kit (Tiangen Biochemical Technology Co., Ltd: DP201101X). The quantity and quality of RNA was determined by an ultramicroscopic spectrophotometer (Thermo Fisher, NanoDrop One). We used a reverse transcription kit to synthesize cDNA and diluted 100 times with ddH$_2$O as templates for subsequent RT-qPCR experiments. Based on the coding sequences of *BnaCOL* genes, specific primers were designed using online website qPCR Primer Database[13]. All *BnaCOL* genes primers were listed in the **Supplementary Table 1**. SYBR® Premix Ex Taq™ (TaKaRa) was used for the real-time quantitative experiment. In this experiment, three biological replicates were collected and the samples without photoperiod treatment were

---

[3]http://smart.embl.de/

[4]https://www.ncbi.nlm.nih.gov/cdd/

[5]http://web.expasy.org/protparam/

[6]https://www.genscript.com/wolf-psort.html

[7]http://gsds.gao-lab.org/

[8]http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index

[9]http://bioinformatics.psb.ugent.be/webtools/plantcare/html

[10]http://eggnog-mapper.embl.de/

[11]http://yanglab.hzau.edu.cn

[12]http://www.heatmapper.ca

[13]https://biodb.swu.edu.cn/qprimerdb/

TABLE 1 | The position and molecular information of _COL_ gene family in _B. napus_.

Gene name	Gene ID	Chromosomes position	CDS (bp)	Protein			Subcellular localization prediction
				Length (bp)	MW (kDa)	pI	
BnaCOL1	BnaA01G0416100ZS	36950508 – 36951440 +	932	310	34.32	5.45	Nuclear
BnaCOL2	BnaA02G0061400ZS	3266686 – 3267639 –	953	317	35.45	6.83	Chloroplast
BnaCOL3	BnaA02G0116400ZS	6119846 – 6121295 –	1,027	342	37.59	5.49	Chloroplast
BnaCOL4	BnaA02G0174500ZS	10523483 – 10524893 –	1,195	398	45.13	5.39	Mitochondria
BnaCOL5	BnaA02G0360300ZS	31881593 – 31883093 –	1,050	350	38.77	5.59	Nuclear
BnaCOL6	BnaA04G0165100ZS	17420919 – 17421857 +	868	289	31.51	7.99	Chloroplast
BnaCOL7	BnaA05G0182600ZS	12667807 – 12668866 +	943	314	37.01	5.95	Nuclear
BnaCOL8	BnaA06G0333200ZS	41587921 – 41589471 +	1,051	350	37.87	5.69	Cytoplasm
BnaCOL9	BnaA06G0365400ZS	43546693 – 43548134 +	1,073	358	39.4	5.85	Nuclear
BnaCOL10	BnaA07G0096900ZS	14681867 – 14683344 +	924	308	34.03	5.8	Nuclear
BnaCOL11	BnaA07G0106200ZS	15211667 – 15213016 –	1,225	408	45.99	5.51	Chloroplast
BnaCOL12	BnaA07G0275200ZS	25743942 – 25745385 +	1,207	402	45.62	5.91	Nuclear
BnaCOL13	BnaA09G0066400ZS	3989150 – 3991166 –	1,039	346	37.39	5.77	Cytoplasm
BnaCOL14	BnaA09G0437200ZS	49558950 – 49560928 –	1,243	414	46.5	5.08	Nuclear
BnaCOL15	BnaA10G0134500ZS	17803925 – 17805427 +	1,185	353	38.78	6.21	Chloroplast
BnaCOL16	BnaA10G0206100ZS	22123527 – 22124540 +	1,013	337	37.94	6.57	Nuclear
BnaCOL17	BnaA10G0206200ZS	22132539 – 22133811 +	1,090	363	41.25	6.73	Nuclear
BnaCOL18	BnaC01G0499200ZS	56776905 – 56777870 –	965	321	35.75	5.58	Nuclear
BnaCOL19	BnaC02G0071200ZS	4363404 – 4364369 –	965	321	35.88	7.89	Chloroplast
BnaCOL20	BnaC02T0142900ZS	10195224 – 10196603 –	1,027	342	37.48	5.63	Chloroplast
BnaCOL21	BnaC02G0484200ZS	58992489 – 58993974 –	1,041	347	38.38	5.5	Nuclear
BnaCOL22	BnaC03G0630900ZS	60651107 – 60652647 +	1,168	389	43.39	6.49	Nuclear
BnaCOL23	BnaC04G0461800ZS	58868888 – 58869910 +	940	313	33.61	7.52	Chloroplast
BnaCOL24	BnaC05G0226300ZS	17178698 – 17180607 +	1,243	414	46.36	5.12	Chloroplast
BnaCOL25	BnaC05G0309900ZS	28672694 – 28673760 +	943	314	36.99	6.2	Nuclear
BnaCOL26	BnaC06G0312000ZS	41664896 – 41666349 +	1,216	405	46.01	6.02	Chloroplast
BnaCOL27	BnaC07G0158900ZS	28546979 – 28548324 –	1,222	407	45.8	5.77	Nuclear
BnaCOL28	BnaC07G0159900ZS	28645926 – 28647275 –	1,225	408	45.83	5.45	Chloroplast
BnaCOL29	BnaC07G0327300ZS	46665480 – 46667263 –	1,076	359	39.62	5.85	Nuclear
BnaCOL30	BnaC07G0361000ZS	48935280 – 48936860 –	1,051	350	37.74	5.46	Cytoplasm
BnaCOL31	BnaC09G0055900ZS	3580725 – 3582245 –	1,033	344	37.34	6.27	Cytoplasm
BnaCOL32	BnaC09G0505400ZS	60998624 – 60999625 +	1,001	333	37.53	6.51	Nuclear
BnaCOL33	BnaC09G0505500ZS	61005355 – 61006629 +	1,099	366	41.71	8.10	Nuclear

MW, molecular weight; pI, isoelectric point.

used as controls. The gene relative expression analysis refers to the $2^{-\Delta\Delta Ct}$ method.
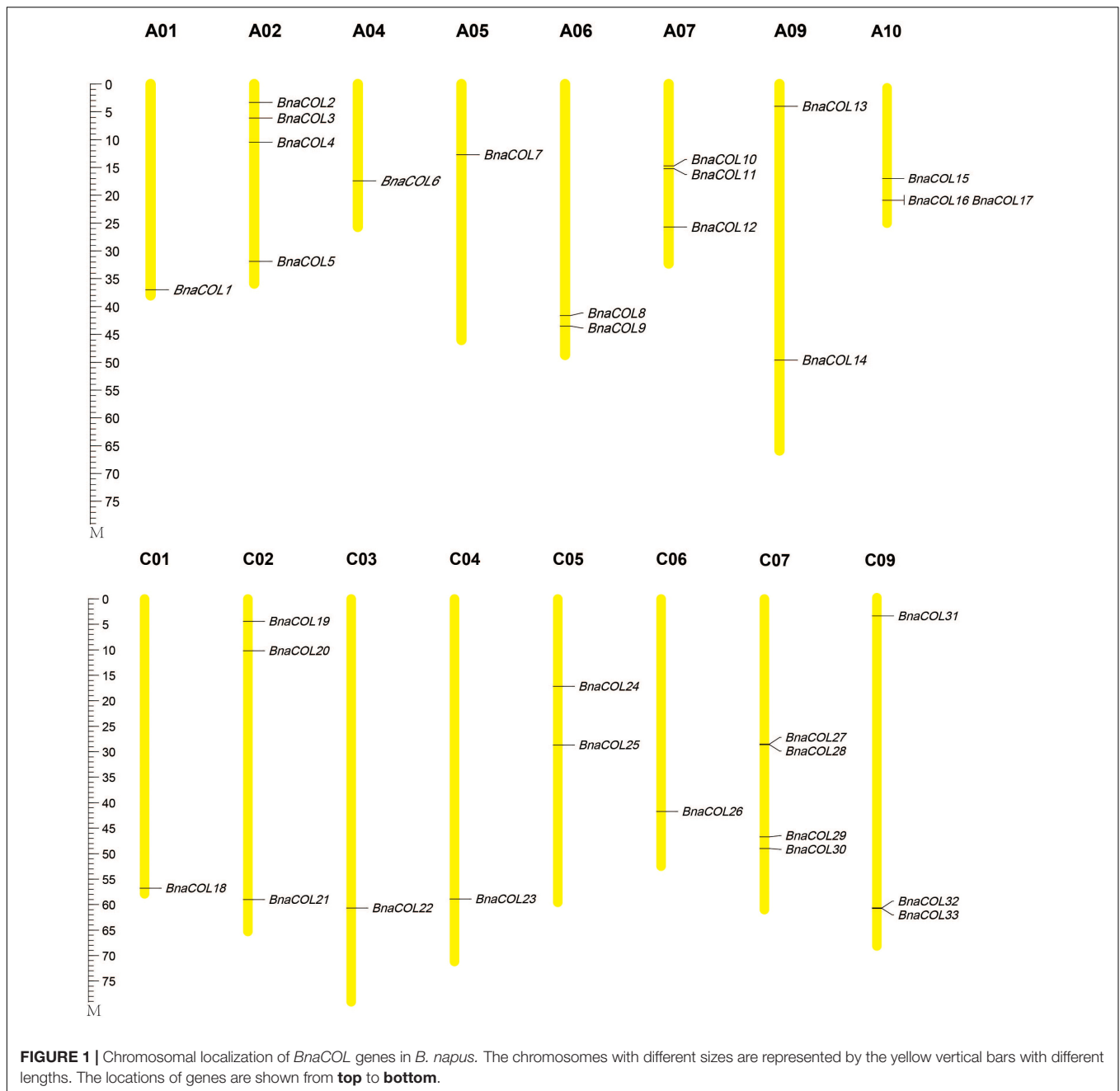
RESULTS

Identification of _CO-Like_ Transcription Factor Family in Rapeseed

We have identified 33 _COL_ genes in _B. napus_ and named them _BnaCOL1-BnaCOL33_ (**Table 1**). Subsequently, the physical and chemical properties of all members were analyzed and predicted. The lengths of the proteins encoded by _BnaCOL_ genes varied from 289 to 414 amino acids, the MW ranged from 31.51 to 46.5 kDa, and the PI ranged from 5.08 to 8.10. The other information about all BnaCOL proteins were list in **Table 1**, including subcellular location prediction and coding sequence length.

Chromosome Location and Phylogenetic Analysis

Furthermore, the 33 _BnaCOL_ genes were mapped on chromosomes (**Figure 1**). These 33 _COL_ genes were distributed unevenly across 16 chromosomes in rapeseed genome and no gene distributed on chromosome A03, A08, and C08. There was only one gene located on chromosome A01, A04, A05, C01, C03, C04, and C06; and two genes on chromosome A06, A09, C05; three genes on chromosome A07, A10, C02, and C09; four genes on chromosome A02 and C07.

To gain a better understanding of the evolutionary relationship between the _COL_ genes of different species, we constructed a phylogenetic tree using 137 _COL_ proteins from seven species, including _Arabidopsis_, _B. oleracea_, _B. rapa_, _B. nigra_, rice, radish, maize (Protein sequences are shown in **Supplementary Table 2**). As shown in **Figure 2A**, these

FIGURE 1 | Chromosomal localization of *BnaCOL* genes in *B. napus*. The chromosomes with different sizes are represented by the yellow vertical bars with different lengths. The locations of genes are shown from **top** to **bottom**.
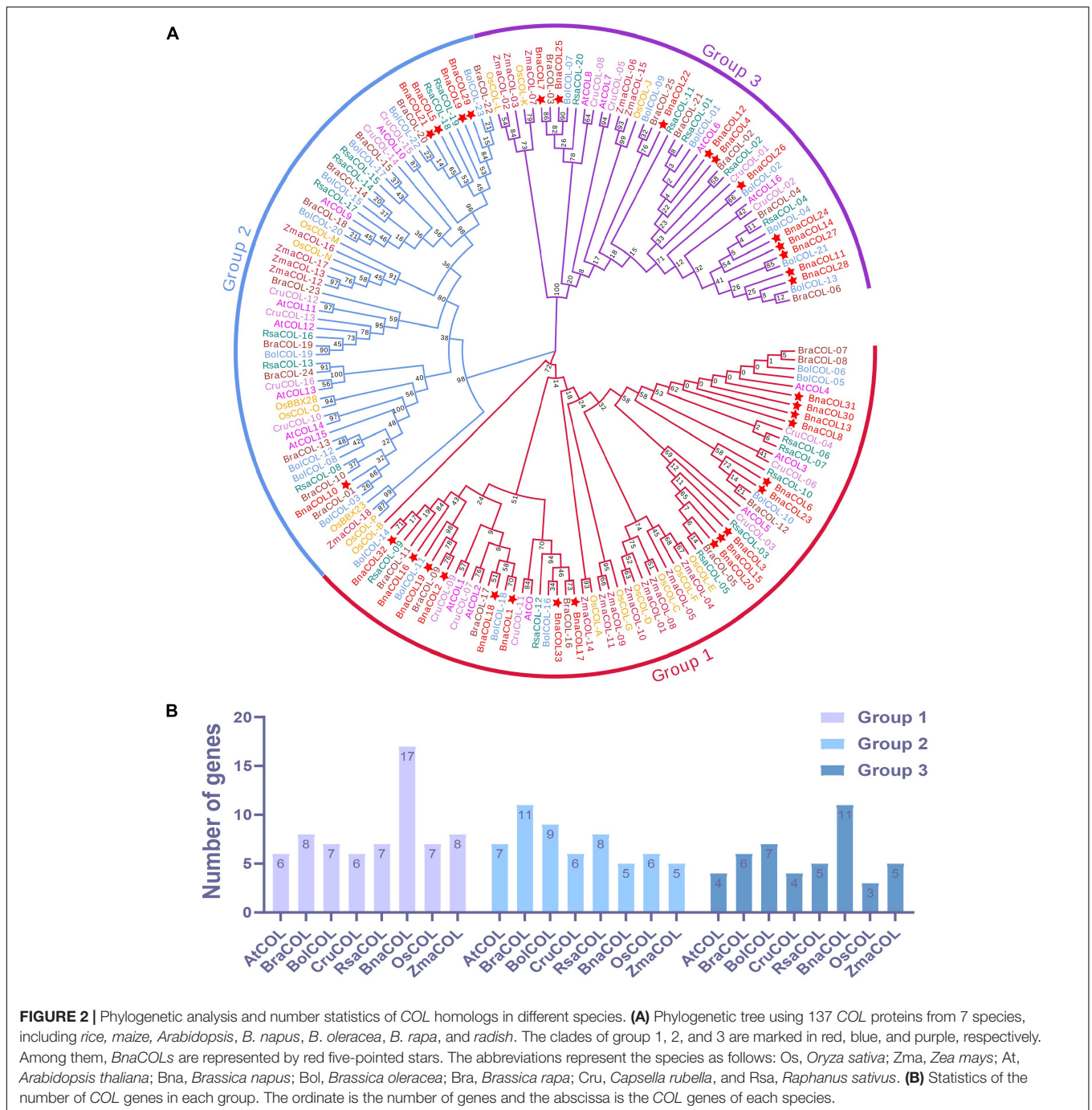
COL genes were classified into three groups. The first group consisting of seven species contained the most *COL* members, while the third group had the least numbers of *COL* genes. For the *BnaCOL* genes, there were 17, 5, and 11 members clustered in to the group 1, 2, and 3, respectively (marked with asterisks in **Figure 2A**). The *BnaCOL* members which were closely grouped may come from a common origin and have similar functions.

Subsequently, we counted the number of *COL* genes of each species in each group (**Figure 2B**). Based on the number of *COL* genes in *A. thaliana*, only about one copy of the *COL* genes were retained in each group of species. It should be pointed out that

the *COL* genes of *B. napus* in Group 1 and Group 3 retain about 3 homologous copies.

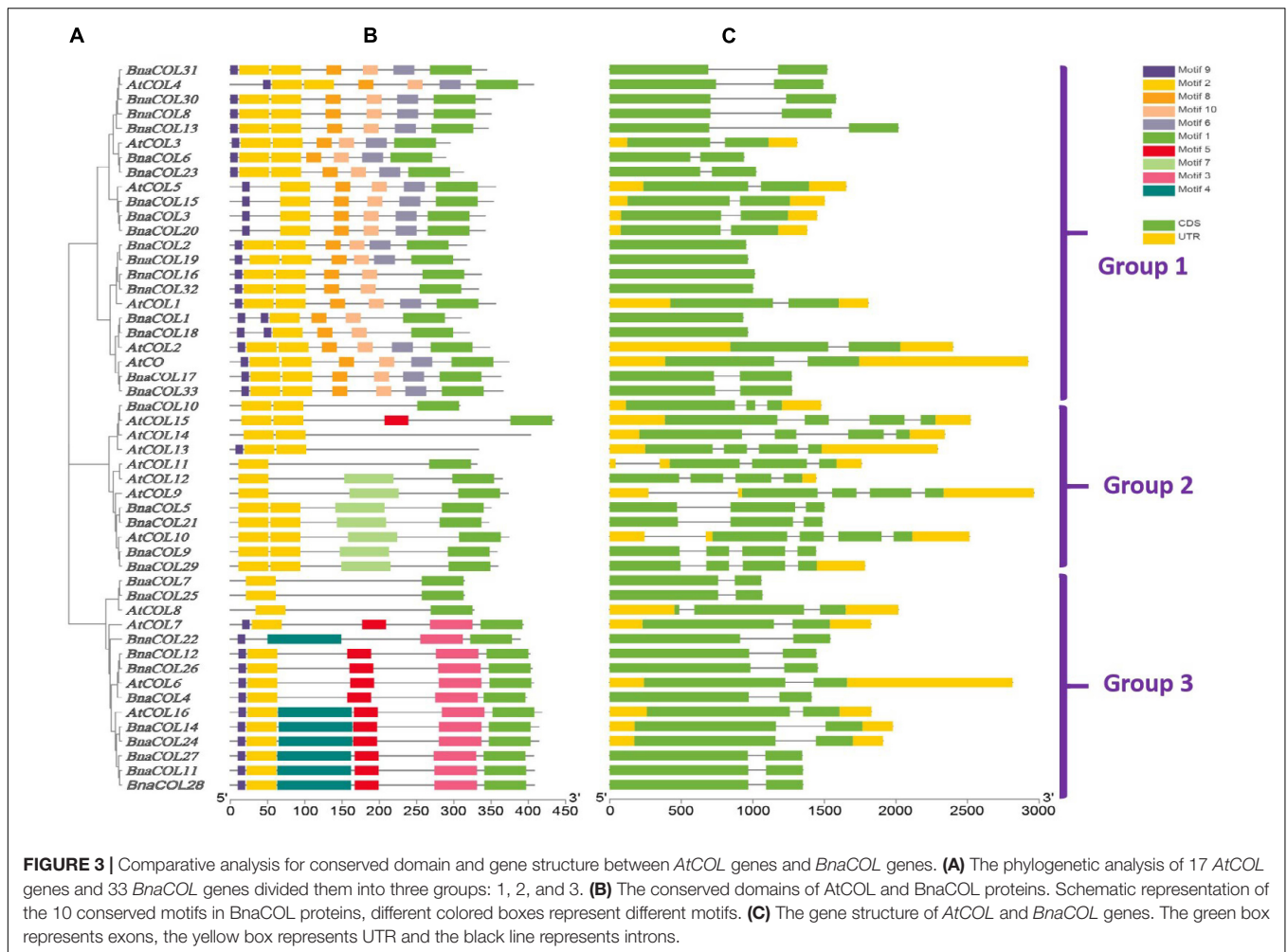## Protein Conservative Domain and Gene Structure Analysis

To investigate the structural diversity of *BnaCOL* genes, we constructed a phylogenetic tree using 17 AtCOL protein sequences from *A. thaliana* and 33 BnaCOL protein sequences from *B. napus*. All of the *COL* genes were classified into three groups: 1, 2, and 3 (**Figure 3A**). And then their protein conserved domains and gene structure were further analyzed.

FIGURE 2 | Phylogenetic analysis and number statistics of *COL* homologs in different species. **(A)** Phylogenetic tree using 137 *COL* proteins from 7 species, including *rice, maize, Arabidopsis*, *B. napus*, *B. oleracea*, *B. rapa*, and *radish*. The clades of group 1, 2, and 3 are marked in red, blue, and purple, respectively. Among them, *BnaCOLs* are represented by red five-pointed stars. The abbreviations represent the species as follows: Os, *Oryza sativa*; Zma, *Zea mays*; At, *Arabidopsis thaliana*; Bna, *Brassica napus*; Bol, *Brassica oleracea*; Bra, *Brassica rapa*; Cru, *Capsella rubella*, and Rsa, *Raphanus sativus*. **(B)** Statistics of the number of *COL* genes in each group. The ordinate is the number of genes and the abscissa is the *COL* genes of each species.

The protein conserved domain analysis revealed a total of 10 different conservative motifs (**Figure 3B**). In general, all members contained motif 1 (CCT domain) and motif 2 (B-box domain), indicating that CCT domain and B-box domain are highly conserved in *BnaCOL* genes. Besides, similar conserved motifs were found in members of the same group. For example, all members of group 1 contained six motifs: motif 9, motif 2, motif 8, motif 10, motif 6, and motif 1 and the distribution and length of these motifs were consistent. Furthermore, most members of group 2 contained three motifs, among which motif

7 only existed in group 2, while the most *BnaCOLs* in group 3 contain six motifs, among which motif 5, motif 3 and motif 4 are unique to members of group 3. However, there were slight differences in the number and distribution of *COL* motifs in different groups. In group 2, four members, i.e., *BnaCOL10, AtCOL13, AtCOL14*, and *AtCOL15* contained conserved motifs different from other members. In group 3, *BnaCOL7, BnaCOL25*, and *AtCOL8* had three fewer motifs than other members.

The gene structure analysis showed that the *BnaCOL* genes in the same group usually had similar exons and introns

FIGURE 3 | Comparative analysis for conserved domain and gene structure between *AtCOL* genes and *BnaCOL* genes. **(A)** The phylogenetic analysis of 17 *AtCOL* genes and 33 *BnaCOL* genes divided them into three groups: 1, 2, and 3. **(B)** The conserved domains of AtCOL and BnaCOL proteins. Schematic representation of the 10 conserved motifs in BnaCOL proteins, different colored boxes represent different motifs. **(C)** The gene structure of *AtCOL* and *BnaCOL* genes. The green box represents exons, the yellow box represents UTR and the black line represents introns.

(**Figure 3C**). Both group 1 and 3 contained two exons and one intron. But there was a little difference in the distribution and quantity of exons and introns in group 2.

Multi Sequence Alignment and Three-Dimensional Structure Prediction of Protein
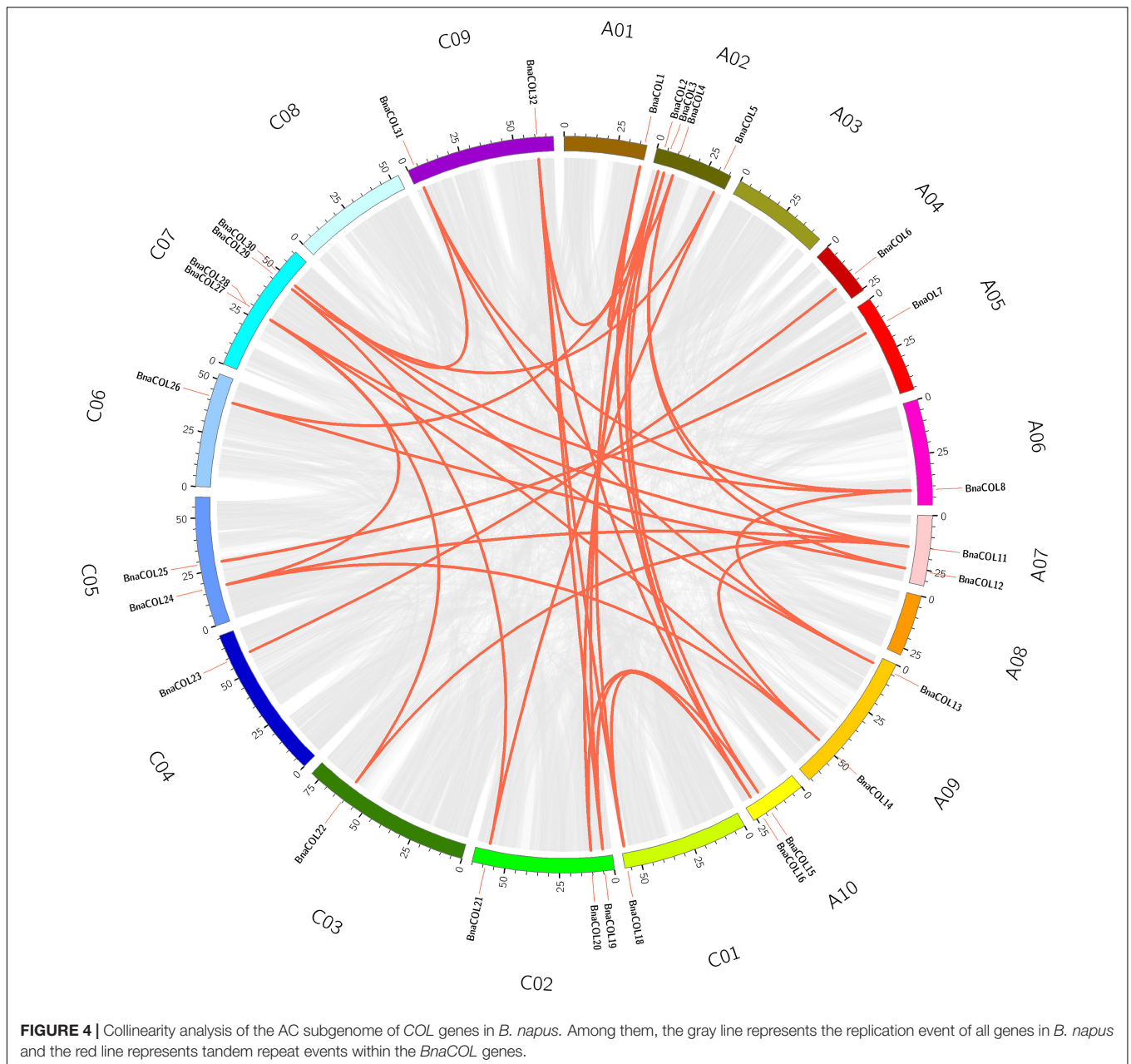
To elucidate the structural characteristics of BnaCOL proteins, we carried out multiple sequence alignment (**Supplementary Figure 1**) and three-dimensional structure prediction analysis (**Supplementary Figure 2**). On the basis of these results we concluded that these proteins have highly conserved CCT and B-box 1 domains, but the sequence of B-box 2 is slightly different.

Then we further predicted the three-dimensional structure of the B-box domain, the results are in good consistent with previous research results (Li et al., 2020). We divided the BnaCOL proteins into three groups according to their genetic relationship. Most of the B-box structure is similar. It is worth noting that group C only contains the predicted B-box 1 domain but not the B-box 2 domain.

Collinearity Analysis Within *Brassica napus* and Among Different Species

Genome wide replication analysis is of great importance for the origin, evolution and genome expansion of species. We hence analyzed the *COL* gene family replication events in *B. napus* to understand the causes of *BnaCOL* genes replication events. The results showed that 37 pairs of large fragment repeat genes were detected (**Figure 4**) and fragment repeats were found on 17 chromosomes except for A03 and C08. These results indicated that large fragment replication may be a major driving force for the amplification and evolution of *COL* genes in *B. napus* genome.
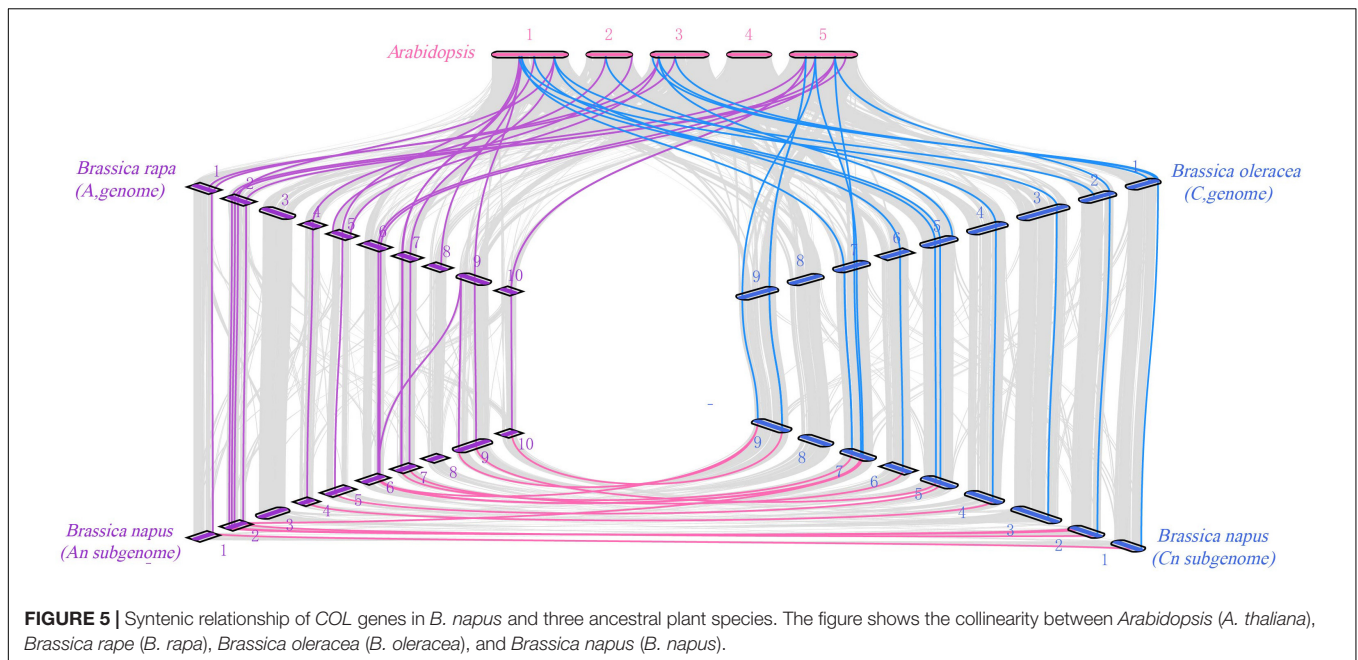
In order to trace the evolutionary process of the *COL* gene family in *Brassica*, we analyzed the homologous relationship among *Arabidopsis*, *B. napus* (A and C subgenomes), *B. rapa* (A genome), and *B. oleracea* (C genome) (**Figure 5**). The collinearity analysis showed that there were a large number of orthologous *COL* genes in *Arabidopsis*, *B. rapa*, *B. oleracea*, and *B. napus*. There were 19 pairs of genes in *Arabidopsis* and *B. rapa* that showed collinearity, 13 *B. rapa COL* genes had homologous genes in *Arabidopsis*, among which 6 were multi-copy genes and 7 were single-copy genes. In addition, *B. rapa* lacked homologous

FIGURE 4 | Collinearity analysis of the AC subgenome of *COL* genes in *B. napus.* Among them, the gray line represents the replication event of all genes in *B. napus* and the red line represents tandem repeat events within the *BnaCOL* genes.

genes of *AtCOL2, AtCOL7, AtCOL11*, and *AtCOL14,* which indicated that gene loss happened in *B. rapa* during evolution. Moreover, 17 pairs of genes in *Arabidopsis* and *B. oleracea* showed collinearity, and 12 *B. oleracea COL* genes had homologous genes in *Arabidopsis* while homologous genes of *AtCOL5, AtCOL7, AtCOL11, AtCOL13*, and *AtCOL14* were not found. The A and C subgenomes of *B. napus* were mainly collinear with the corresponding diploid *B. rapa* and *B. oleracea*. The A genome of *B. napus* and *B. rapa* had 16 homologous gene pairs, while 14 homologous gene pairs were found between the C genome of *B. napus* and *B. oleracea*. For the evolution of *COL* gene family, although gene loss occurs, the vast majority of *COL* genes remain intact in *B. napus*.

Cis-Acting Element and Functional Annotation Analysis

The *cis*-acting element is the binding site of transcriptional regulators and regulates gene transcription. In order to investigate the potential function of the *BnaCOL* genes, we analyzed the *cis*-acting elements of the upstream sequence of the *BnaCOL* promoters at 1,500 bp and excluded the elements with unknown function and the general transcriptional regulatory elements (**Table 2** and **Figure 6**). These *cis*-acting elements can be broadly classified into four categories, which involved in light response, hormonal response, growth regulation, and abiotic-stress response. Among them, the components involved in the light reaction include G-box, GATA-motif, Box4, TCT-motif,

FIGURE 5 | Syntenic relationship of *COL* genes in *B. napus* and three ancestral plant species. The figure shows the collinearity between *Arabidopsis* (*A. thaliana*), *Brassica rape* (*B. rapa*), *Brassica oleracea* (*B. oleracea*), and *Brassica napus* (*B. napus*).

ATCT-motif, Circadian, AAAC-motif, AE-box, TCCC-motif, GT1-motif, 3-AF1 binding site and MRE. Hormone response elements include TCA-element, TGACG-motif and CGTCA-motif, ABRE. In addition, several stress response elements such as TC-rich repeats, LTR, MBS were observed. These results showed that most of the *BnaCOL* genes had photoresponsive elements indicating that *BnaCOL* genes may played a critical role in the regulation of photoreactivity.

Taking the above observations in account, we performed GO annotation and enrichment analysis of *BnaCOL* genes to gain a better understanding of their function. The analysis results mainly included three aspects: biological process (BP), molecular function (MF) and cellular component (CC) (**Supplementary Table 4** and **Supplementary Figure 3**). In the biological process (BP), most genes were annotated in light signal response and transmission, photoperiod response, flowering regulation, circadian rhythm, etc. This is consistent with the observation from the *cis*-acting element. In the molecular function (MF), a total of 11 highly enriched items were detected, including the combination of DNA, protein and organic compounds and transcription regulator activity. Likewise, in the cellular component (CC), most gene annotations were located on the nucleus and organelles. This indicates a good consistency between the prediction of subcellular location and GO enrichment analysis.

Tissue-Specific Expression Pattern of *BnaCOL* Genes

To further study the expression patterns of the *BnaCOL* genes in different tissues of rapeseed, we used the online website (BnTIR: *Brassica napus* transcriptome information resource) to download the rapeseed genome-wide transcription data of different tissues. As show in **Figure 7** and **Supplementary Table 5**, all

BnaCOL genes showed different expression characteristics in various tissues, which indicated that *BnaCOL* genes were usually not tissue-specific genes. A 81.8% of *BnaCOL* genes showed high expression levels in leaves and sepals, while *BnaCOL5, BnaCOL9, BnaCOL10, BnaCOL21*, and *BnaCOL29* were abundantly expressed in pollen and flower buds. Nevertheless, *BnaCOL6, BnaCOL7, BnaCOL13, BnaCOL23*, and *BnaCOL31* were not only highly expressed in leaves and sepals, but also highly detected in siliques. *BnaCOL6* and *BnaCOL23* expressed the highest levels in siliques. In particular, the expression of *BnaCOL22* was lower in other tissues, but highest in seeds. On the basis of these results we concluded that the *BnaCOL* gene family were critical for all stages of the development of rapeseed individuals and some members with similar expression characteristics may perform similar functions.

Diurnal Rhythm of Expression of *BnaCOL* Genes

Previously, we analyzed the *cis*-acting elements in the upstream sequence of the *BnaCOL* promoters and found that most *BnaCOL* genes have light-responsive elements, indicating that they may involve in photoperiod regulation. To further identify the possible function, we selected nine homologous genes that regulate flowering in *Arabidopsis* (*BnaCOL3, BnaCOL5, BnaCOL11, BnaCOL12, BnaCOL15, BnaCOL16, BnaCOL23, BnaCOL30, BnaCOL33*) and tested the circadian expression profile of these nine genes within 24 h (**Figures 8, 9**).

The circadian expression pattern of *BnaCOL* genes under LD illumination showed four types: (i) The expression of *BnaCOL3, BnaCOL11, BnaCOL15*, and *BnaCOL23* increased slowly during illumination, peaked at 12 h and then rapidly decreased to a lower level. (ii) While *BnaCOL30* and *BnaCOL33* had similar expression patterns, their expression levels presented a stepwise

TABLE 2 | The *cis*-elements have been identified in more than three *BnaCOL* genes.

Site name	Sequence	Function of the *cis*-elements
G-Box	CACGTG	*cis*-acting regulatory element involved in light responsiveness
GATA-motif	AAGATAAGATT	part of a light responsive element
TCA-element	CCATCTTTTT	*cis*-acting element involved in salicylic acid responsiveness
Box 4	ATTAAT	part of a conserved DNA module involved in light responsiveness
TCT-motif	TCTTAC	part of a light responsive element
ATCT-motif	AATCTAATCC	part of a conserved DNA module involved in light responsiveness
circadian	CAAAGATATC	*cis*-acting regulatory element involved in circadian control
TC-rich repeats	ATTCTCTAAC	*cis*-acting element involved in defense and stress responsiveness
LTR	CCGAAA	*cis*-acting element involved in low-temperature responsiveness
AAAC-motif	CAATCAAAACCT	light responsive element
AE-box	AGAAACAA	part of a module for light response
TCCC-motif	TCTCCCT	part of a light responsive element
GT1-motif	GGTTAAT	light responsive element
TGACG-motif	TGACG	*cis*-acting regulatory element involved in the MeJA-responsiveness
CCAAT-box	CAACGG	MYBHv1 binding site
CGTCA-motif	CGTCA	*cis*-acting regulatory element involved in the MeJA-responsiveness
ABRE	ACGTG	*cis*-acting element involved in the abscisic acid responsiveness
3-AF1 binding site	TAAGAGAGGAA	light responsive element
MRE	AACCTAA	MYB binding site involved in light responsiveness
MBS	CAACTG	MYB binding site involved in drought-inducibility

increase pattern. Whether it is day or night, they reach their peak in 20 h or 24 h. (iii) On the contrary, *BnaCOL5* was up-regulated before 4 h, then gradually declined. (iv) In particular, the expression level of *BnaCOL16* showed light inhibition and dark induction. The expression reached its peak at 0 h and decreased rapidly to a lower level under light conditions, then slowly increased to dark. However, *BnaCOL12* had no significant difference in expression level under LD illumination.

The circadian expression pattern of *BnaCOL* genes under SD illumination showed three types: (i) The expression level of *BnaCOL3, BnaCOL16*, and *BnaCOL30* gradually decreased during the light period, decreased to a minimum at 8 h or 12 h and then slowly rised, showing U-shaped curve. (ii) However, the expression levels of *BnaCOL5, BnaCOL11, BnaCOL12*, and *BnaCOL23* increased slowly from 0 h, reached a peak at 12 h, whereafter gradually dropped down. (iii) It was interesting to find that the expression level of *BnaCOL15* was lower during the light period and reached the highest at 20 and 24 h under dark conditions, indicating that the dark treatment can activate the expression of *BnaCOL15*.

In general, most *BnaCOL* genes exhibited different diurnal expression patterns, indicating that these *COL* genes were involved in the photoperiod pathway. However, the expression patterns of *BnaCOL12* and *BnaCOL16* were similar under LD or SD conditions, indicating that these two genes may not be affected by photoperiod.
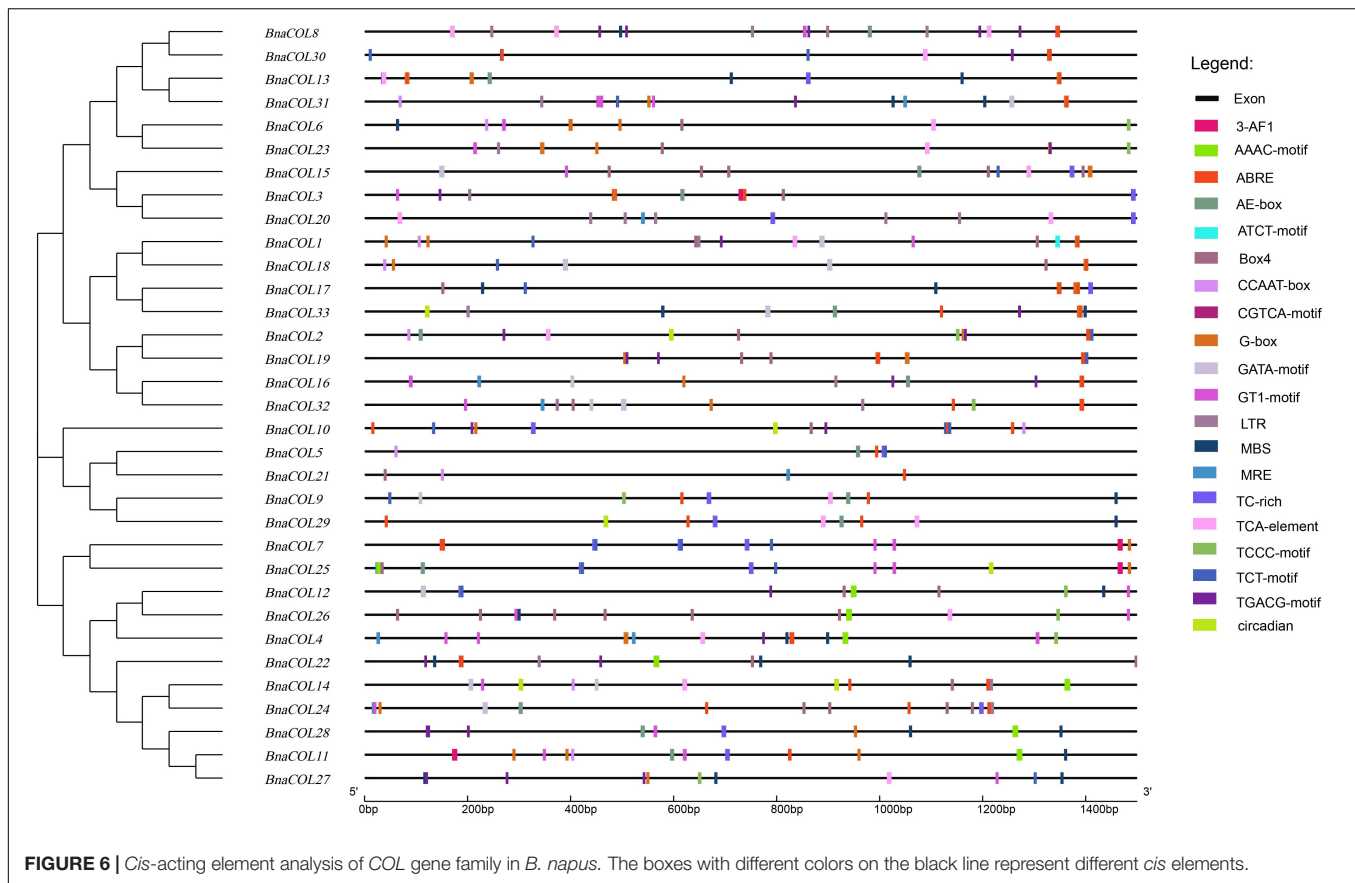
DISCUSSION

In this study, we conducted a biological analysis of the *COL* gene family of *B. napus* and related species, including their chromosomal location, phylogenetic analysis, gene structure, protein conserved domain analysis and three-dimensional structure prediction of protein. The results showed that 33 *BnaCOL* family members were unevenly distributed on 16 chromosomes of rapeseed. Based on phylogenetic analysis, these members were divided into three groups. And most of the genes in the same subgroup had similar gene structure, protein conserved domains and three-dimensional protein structure, which reflects the conservation of the *BnaCOL* gene family.

The Retention and Deletion of *CONSTANS-LIKE* Genes in *Brassica* During Evolution

Most angiosperms (including monocots and dicots) have experienced whole-genome duplications (WGDs). In the dicotyledon *A. thaliana,* it has experienced three WGDS events: two ancient tetraploid events (β, α) and one ancient hexaploid (γ) event (Bowers et al., 2003; Tang et al., 2008). And previous studies have shown that the ancestral genomes of Brassica species (similar in structure to *A. thaliana*) have undergone genome-wide triploid replication events, indicating that they evolved from a common hexaploid ancestor (Lysak et al., 2005; Town et al., 2006). *B. napus* (A and C genomes) is the amphidiploid species formed by a cross between *B. rapa* (A genome) and *B. oleracea* (C genome). As such, the rape genome contains six times the ancestral genome (Parkin et al., 2002). During the evolution of *Brassica*, it remains question to be explored whether the retention and deletion of flowering and photoperiod genes like *COL* accompanying with the evolution of *Brassica* genome.

In the process of polyploidization, copy number variation is an important way of evolution. Taking the number of *COL* genes in *Arabidopsis* as a reference, the number of *COL* genes in *B. oleracea, B. rapa, B. nigra,* radish, rice and maize in each group only retained roughly one copy. We speculate that the *COL* genes have basically undergone recombination and fusion during the evolution process. It should be noted that the retention rate of *BnaCOL* genes in the three groups were quite different. In the first group, 47.2% of the *BnaCOL* genes were retained. In the second group, only 11.9% of the *BnaCOL* genes remained. The third group, the *BnaCOL* genes retention rate was 45.8%. On the basis of our findings, it can be concluded that the *AtCOL* genes of the first and third groups had about three orthologous genes in *B. napus*, while the *AtCOL* genes of the second group only retained one copy in *B. napus,* indicating that the *COL* gene family shrank during the diversification of various species. These
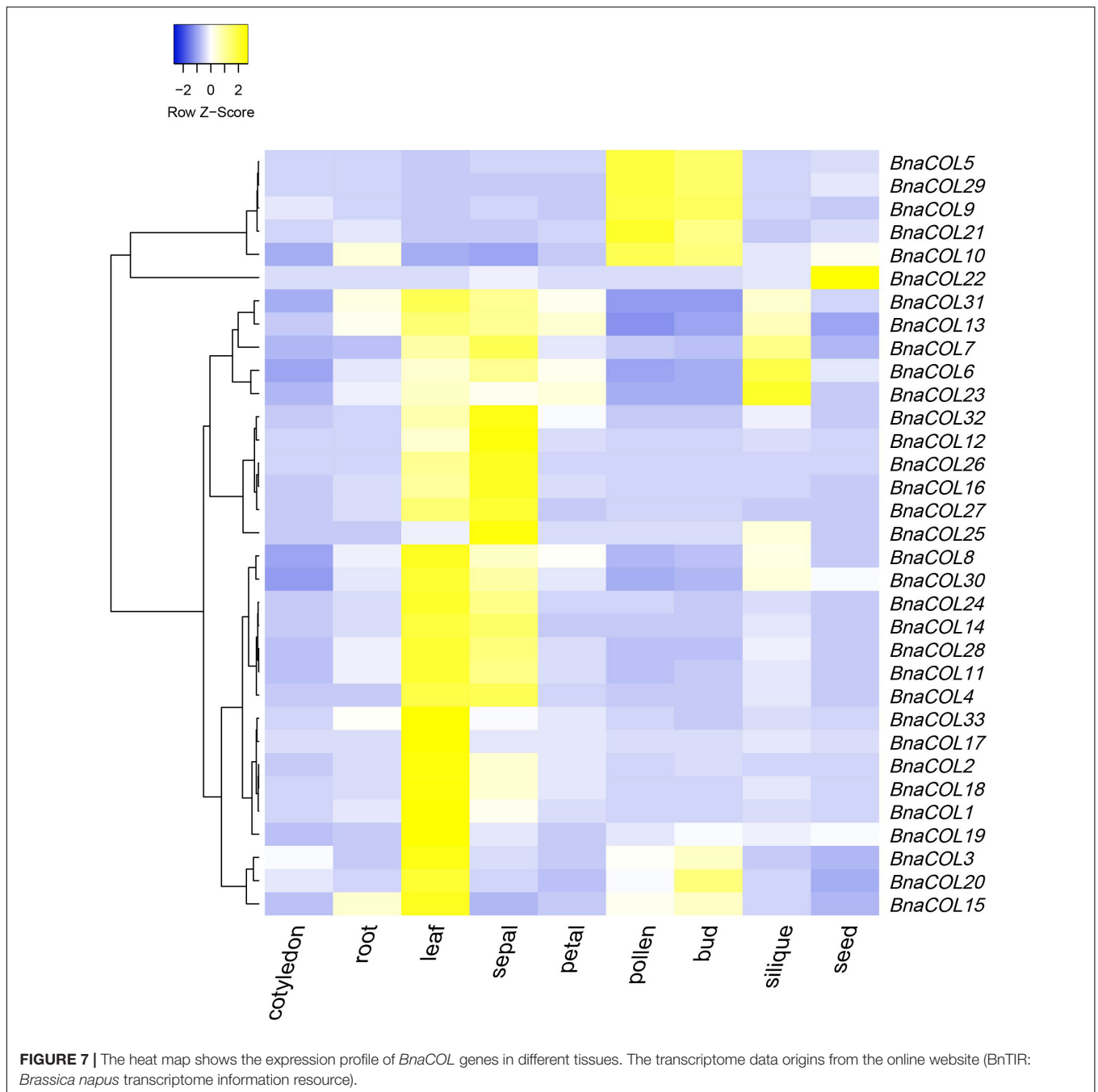
FIGURE 6 | *Cis*-acting element analysis of *COL* gene family in *B. napus.* The boxes with different colors on the black line represent different *cis* elements.

missing *COL* genes may be redundant genes, gradually being replaced by other genes with similar functions.

To further explore the evolutionary process of the *COL* gene family in *Brassica*, we analyzed the replication events of the *COL* gene family in *B. napus.* The results revealed a total of 37 pairs of large-segment repetitive genes existed. We also analyzed the homology relationships among *Arabidopsis*, *B. napus*, *B. rapa*, and *B. oleracea*. The results of collinearity analysis showed that there were a large number of homologous *COL* genes in *B. napus*, *Arabidopsis*, *B. rapa*, and *B. oleracea*. However, *B. rapa* lacks homologous genes of *AtCOL2, AtCOL7, AtCOL11*, and *AtCOL14*, while *B. oleracea* lacks homologous genes of *AtCOL5, AtCOL7, AtCOL11, AtCOL13*, and *AtCOL14*, indicating that gene loss occurred in the process of evolution. This results are basically consistent with the previous phylogenetic tree analysis. As mentioned earlier, there should be six homologous copies of each *Arabidopsis* gene in *B. napus* genome. As 17 *COL* genes were identified in *Arabidopsis* (Robson et al., 2001). In theory, there should be 102 *COL* genes in *B. napus* after whole genome replication, but only 33 *BnaCOL* genes were identified in this study, indicating that about 67.7% of them were lost after whole genome replication. We hypothesized that during the evolution of *B. napus*, the *COL* genes may have undergone strict purification and selection, which play a key role in the maintenance of gene number.

Functional Expression Diversity of *BnaCOL* Genes in *Brassica napus*

We analyzed expressional pattern of 33 *BnaCOL* genes in the different tissues using public data. The results showed that they were expressed in different tissues and their expression patterns in different tissues were different. A 81.8% of *BnaCOL* genes were highly transcribed in leaves and sepals and main expression of five genes were found in pollen and buds (*BnaCOL5, BnaCOL9, BnaCOL10, BnaCOL21*, and *BnaCOL29*). *BnaCOL6, BnaCOL7*, and *BnaCOL23* showed higher transcription levels in siliques while *BnaCOL22* only highly expressed in seeds but lower in other tissues. These results indicate that the *BnaCOL* gene family may play a significant role in all stages of rapeseed growth and development. Since 81.8% of *BnaCOL* genes mainly expressed in leaves, they may have a potential key role in leaves to response to environmental factors like light condition. Previous studies had shown that CO initiated the transcriptional expression of FT, which in turn transferred FT proteins from leaf phloem to shoot tip meristem, thereby activate flowering (Shim et al., 2017). Hence, the high expression of *BnaCOL* genes help to initiate FT transcription, consequently promote flowering.
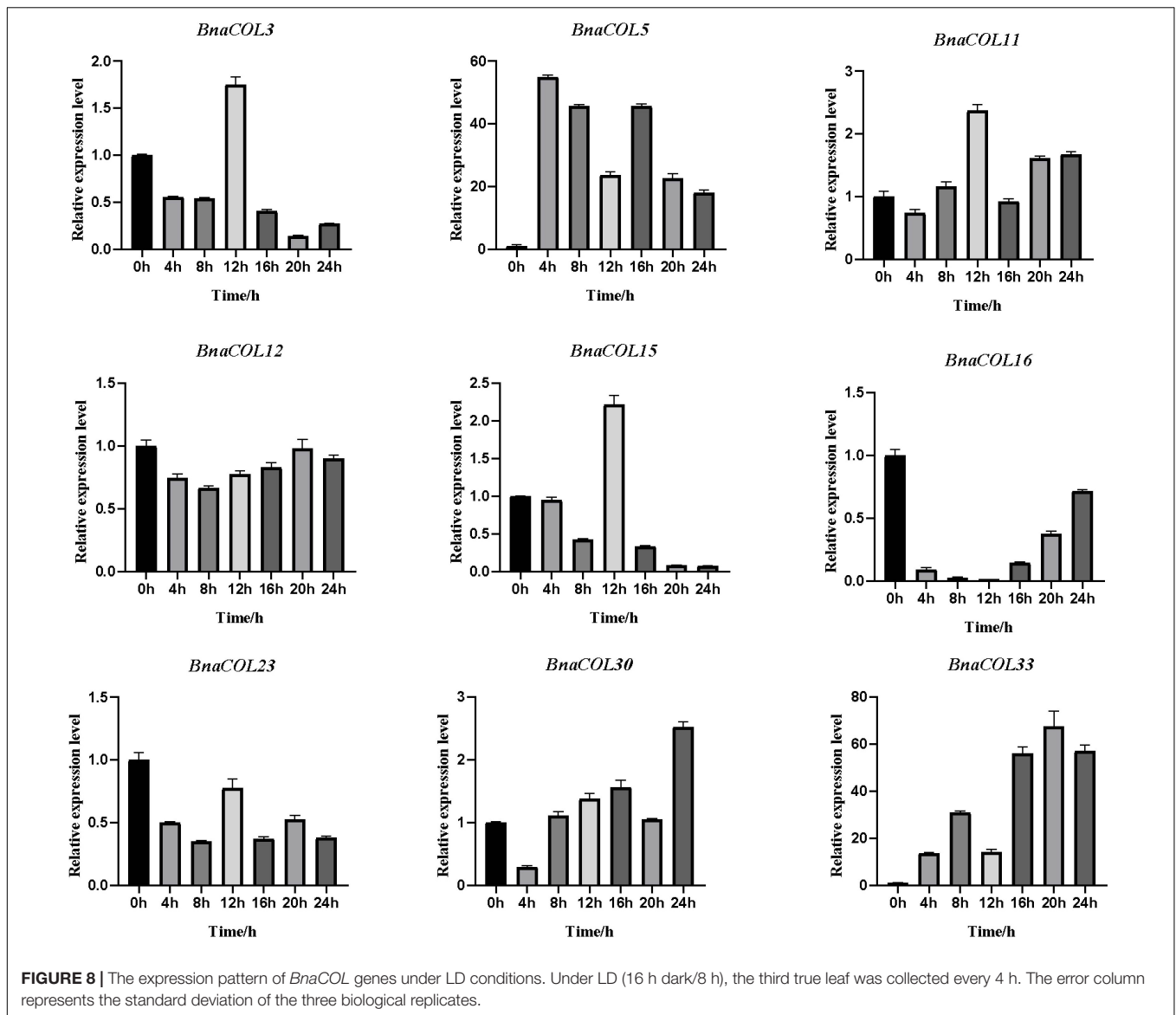
The study results of *COL* genes in *Arabidopsis* provide reference and clue for the *BnaCOL* genes function. Previous studies had identified that overexpression of *AtCOL1* in *Arabidopsis* shorten the circadian cycle (Ledger et al., 2001). And *AtCOL3, AtCOL4, AtCOL5, AtCOL8*, and *AtCOL9* were

FIGURE 7 | The heat map shows the expression profile of *BnaCOL* genes in different tissues. The transcriptome data origins from the online website (BnTIR: *Brassica napus* transcriptome information resource).

all involved in the regulation of flowering under different photoperiod conditions (Cheng and Wang, 2005; Datta et al., 2006; Hassidim et al., 2009; Takase et al., 2012; Steinbach, 2019). Hence, in order to further explore the role of *BnaCOL* genes in the photoperiod pathway, we selected nine *Arabidopsis* homologous genes that regulate flowering time (*BnaCOL3, BnaCOL5, BnaCOL11, BnaCOL12, BnaCOL15, BnaCOL16, BnaCOL23, BnaCOL30,* and *BnaCOL33*) and detected the diurnal expression profiles of these nine genes within 24 h under LD and SD illumination treatments. The RT-qPCR results showed that most members had different daily expression patterns between LD and

SD conditions, revealing the functional difference and divergence of the *BnaCOL* family members in *B. napus*.

The genes of the same evolutionary branch between species are homologous or closely related and their biological functions are roughly the same. Notably, previous studies have shown that *AtCOL1* and *AtCOL4* transcription levels are regulated by circadian rhythms (Ledger et al., 2001; Steinbach, 2019). Our results showed that *BnaCOL16*, as a homologous gene of *AtCOL1*, was expressed at exactly the same level as *AtCOL1* in LD and SD. Therefore, it is speculated that *BnaCOL16* has similar functions to *AtCOL1*. Besides, *AtCOL4* is not only a flowering inhibitor

FIGURE 8 | The expression pattern of *BnaCOL* genes under LD conditions. Under LD (16 h dark/8 h), the third true leaf was collected every 4 h. The error column represents the standard deviation of the three biological replicates.

under LD and SD (Steinbach, 2019), but also participates in ABA and salt stress responses (Min et al., 2015). Our RT-qPCR results showed that *BnaCOL30*, as a homolog of *AtCOL4*, had a circadian rhythm expression profile consistent with that of *AtCOL4*. And the response to abiotic stimuli was also detected in the GO annotation of *BnaCOL30* gene. This implied that *BnaCOL30* may have the same function as *AtCOL4*. In the course of evolution, structurally similar genes may sometimes diverge functionally within or between species. Previous studies have shown that overexpression of *AtCOL5* makes plants early flowering in SD light conditions (Hassidim et al., 2009). In our study, the expression levels of *AtCOL5* and its homologous gene *BnaCOL15* were completely opposite under SD light conditions. The *BnaCOL15* gene may have functional differentiation and its specific function needs to be verified by subsequent experiments. In our results, the expression level of *BnaCOL11* is regulated by the circadian rhythm. But the *AtCOL7* gene, which is closely

related to *BnaCOL11*, has only reported the function of regulating the branching and shading response of *Arabidopsis thaliana* (Wang et al., 2013), as well as linking light perception with auxin (Zhang et al., 2014). In addition, the *BnaCOL* genes exhibit several types of circadian rhythms, suggesting that functional differences in the *BnaCOL* family responsive to multiple aspects of plant development, including the regulation of flowering. Although the *COL* genes are highly conserved among species, it will also undergo functional differentiation in the course of evolution to adapt to environmental changes. The analysis of *cis*-acting elements in the upstream sequence of *BnaCOL* promoters revealed that *BnaCOL* genes contained elements of light response, hormone response, growth regulation and abiotic-stress response. However, only the role of the *BnaCOL* family in the photoperiod pathway was identified in this study, there are many additional unknown functions of the *BnaCOL* gene family required to be explored.

FIGURE 9 | The expression pattern of *BnaCOL* genes under SD conditions. Under SD (8 h light/16 h dark), the third true leaf was collected every 4 h. The error column represents the standard deviation of the three biological replicates.
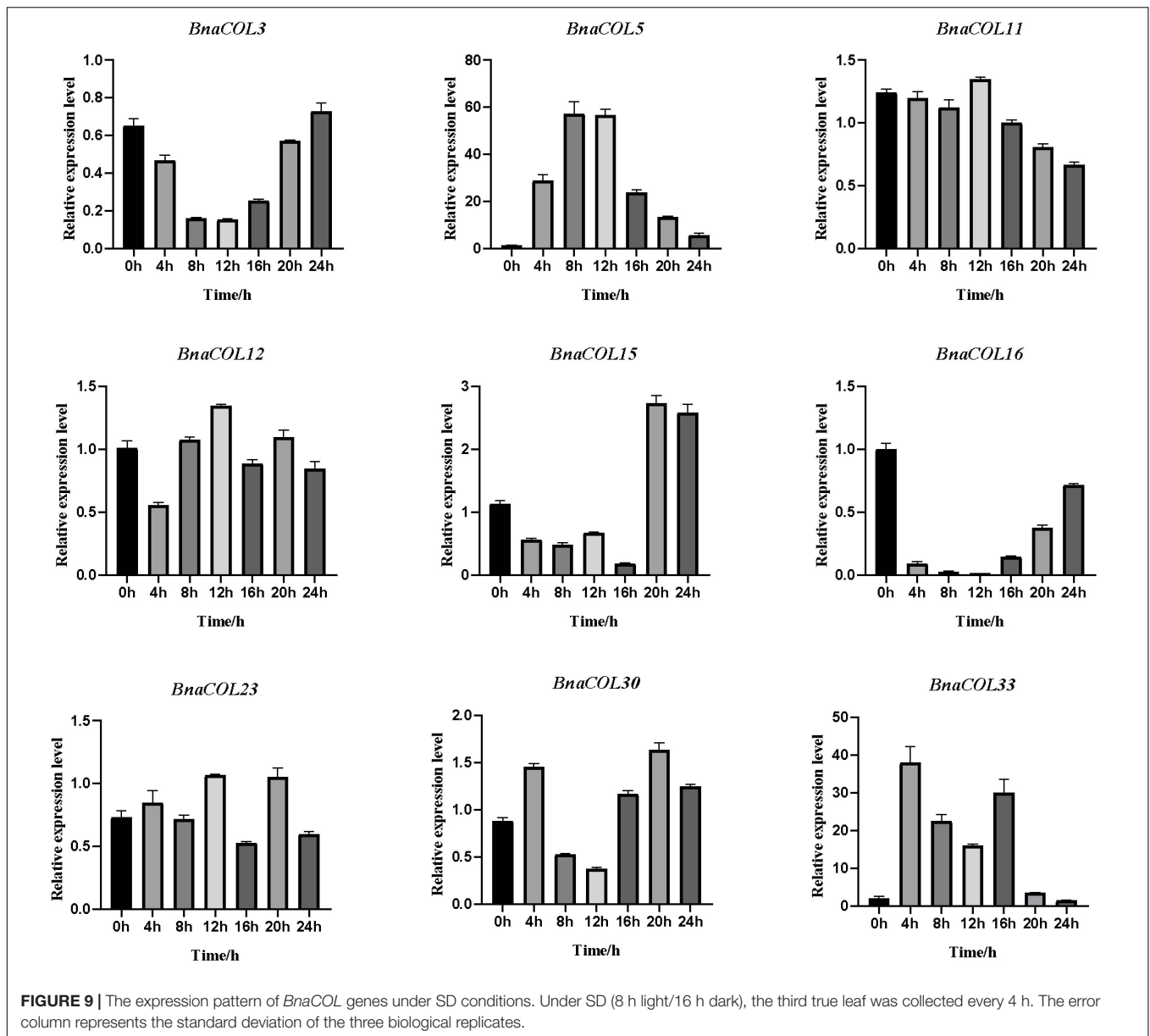
CONCLUSION

In summary, a total of 33 *BnaCOL* genes were identified in *B. napus* and these genes were distributed unevenly on 16 chromosomes. The phylogeny, gene structure, conserved motifs and three-dimensional structure of the COL proteins were analyzed. These genes were classified into three subfamilies and relatively conservative gene structures and motifs were found in the same subfamily. In addition, the *BnaCOLs* promoter region have light-responsive *cis*-elements, as well as a variety of *cis*-acting elements related to hormones and abiotic-stress response. Subsequently, GO annotation and enrichment analysis of *BnaCOL* genes lead us to conclude that most genes are annotated in light signal response and transmission, photoperiod

response, flowering regulation, circadian rhythm, etc. The collinearity analysis found 37 pairs of large-segment repetitive genes in *B. napus*. Based on comparative genomics research, the *COL* genes of *B. napus* had undergone polyploidization and different degrees of loss and expansion. We also analyzed the expression patterns of the *BnaCOL* genes in different tissues of rapeseed, which indicated that the *BnaCOL* gene family were of great importance at various developmental stages in *B. napus*. Besides, we tested the diurnal expression profiles of 9 *BnaCOL* genes under LD and SD conditions. Most members showed different daily expression patterns between LD and SD conditions, revealing the functional differences of the *BnaCOL* family in *B. napus*. In general, this article comprehensively analyzed the conservation and divergence of *BnaCOL* family

genes functions, which provided a biological basis for the further functional identification of *COL* genes in cruciferous plants.

AUTHOR CONTRIBUTIONS

YC, WW, and JL designed the experiments. YC and RZ performed the experiments. YC, QH, and JL analyzed the data. YC wrote the manuscript. All authors reviewed and approved the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.760379/full#supplementary-material

Supplementary Figure 1 | Multiple sequence alignment of AtCOL and BnaCOL proteins.

Supplementary Figure 2 | Three-dimensional structure prediction of B-box domain of BnaCOL proteins. **(A)** Two B-box domains from group I. **(B)** A B-box domain and a second divergent B-box domain from group II. **(C)** One B-box domain from group III. The images are from dark blue to dark red, indicating from the N end to the C end.

Supplementary Figure 3 | GO annotation of *BnaCOL* genes in *B. napus*.

REFERENCES

Borden, K. L. (1998). RING fingers and B-boxes: zinc-binding protein-protein interaction domains. *Biochem. Cell Biol.* 76, 351–358. doi: 10.1139/bcb-76-2-3-351

Bowers, J. E., Chapman, B. A., Rong, J., and Paterson, A. H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422, 433–438. doi: 10.1038/nature01521

Cai, D., Liu, H., Sang, N., and Huang, X. (2017). Identification and characterization of *CONSTANS-like (COL)* gene family in upland cotton (*Gossypium hirsutum* L.). *PLoS One* 12:e0179038. doi: 10.1371/journal.pone.0179038

Cheng, X. F., and Wang, Z. Y. (2005). Overexpression of *COL9*, a *CONSTANS-LIKE* gene, delays flowering by reducing expression of CO and FT in *Arabidopsis thaliana*. *Plant J.* 43, 758–768. doi: 10.1111/j.1365-313X.2005.02491.x

Chia, T. Y., Müller, A., Jung, C., and Mutasa-Göttgens, E. S. (2008). Sugar beet contains a large *CONSTANS-LIKE* gene family including a CO homologue that is independent of the early-bolting (B) gene locus. *J. Exp. Bot.* 59, 2735–2748. doi: 10.1093/jxb/ern129

Dally, N., Eckel, M., Batschauer, A., Höft, N., and Jung, C. (2018). Two *CONSTANS-LIKE* genes jointly control flowering time in beet. *Sci. Rep.* 8:16120. doi: 10.1038/s41598-018-34328-4

Datta, S., Hettiarachchi, G. H., Deng, X. W., and Holm, M. (2006). *Arabidopsis CONSTANS-LIKE3* is a positive regulator of red light signaling and root growth. *Plant Cell* 18, 70–84. doi: 10.1105/tpc.105.038182

Fitter, A. H., and Fitter, R. S. (2002). Rapid changes in flowering time in British plants. *Science* 296, 1689–1691. doi: 10.1126/science.1071617

González-Schain, N. D., and Suárez-López, P. (2008). *CONSTANS* delays flowering and affects tuber yield in potato. *Biol. Plant.* 52:251. doi: 10.1111/j.1365-313X.2012.04909.x

Griffiths, S., Dunford, R. P., Coupland, G., and Laurie, D. A. (2003). The evolution of *CONSTANS-like* gene families in barley, rice, and *Arabidopsis*. *Plant Physiol.* 131, 1855–1867. doi: 10.1104/pp.102.016188

Hassidim, M., Harir, Y., Yakir, E., Kron, I., and Green, R. M. (2009). Over-expression of *CONSTANS-LIKE 5* can induce flowering in short-day grown *Arabidopsis*. *Planta* 230, 481–491. doi: 10.1007/s00425-009-0958-7

Hayama, R., Yokoi, S., Tamaki, S., Yano, M., and Shimamoto, K. (2003). Adaptation of photoperiodic control pathways produces short-day flowering in rice. *Nature* 422, 719–722. doi: 10.1038/nature01549

Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., and Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31, 1296–1297. doi: 10.1093/bioinformatics/btu817

Hu, T., Wei, Q., Wang, W., Hu, H., Mao, W., Zhu, Q., et al. (2018). Genome-wide identification and characterization of *CONSTANS-like* gene family in radish (*Raphanus sativus*). *PLoS One* 13:e0204137. doi: 10.1371/journal.pone.0204137

Kim, S. K., Yun, C. H., Lee, J. H., Jang, Y. H., Park, H. Y., and Kim, J. K. (2008). *OsCO3*, a *CONSTANS-LIKE* gene, controls flowering by negatively regulating the expression of *FT-like* genes under SD conditions in rice. *Planta* 228, 355–365. doi: 10.1007/s00425-008-0742-0

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Ledger, S., Strayer, C., Ashton, F., Kay, S. A., and Putterill, J. (2001). Analysis of the function of two circadian-regulated *CONSTANS-LIKE* genes. *Plant J.* 26, 15–22. doi: 10.1046/j.1365-313x.2001.01003.x

Lee, Y. S., Jeong, D. H., Lee, D. Y., Yi, J., Ryu, C. H., Kim, S. L., et al. (2010). *OsCOL4* is a constitutive flowering repressor upstream of Ehd1 and downstream of OsphyB. *Plant J.* 63, 18–30. doi: 10.1111/j.1365-313X.2010.04226.x

Li, J., Gao, K., Yang, X., Khan, W. U., Guo, B., Guo, T., et al. (2020). Identification and characterization of the *CONSTANS-like* gene family and its expression profiling under light treatment in *Populus*. *Int. J. Biol. Macromol.* 161, 999–1010. doi: 10.1016/j.ijbiomac.2020.06.056

Liu, B., Long, H., Yan, J., Ye, L., Zhang, Q., Chen, H., et al. (2021). A HY5-COL3-COL13 regulatory chain for controlling hypocotyl elongation in *Arabidopsis*. *Plant Cell Environ.* 44, 130–142. doi: 10.1111/pce.13899

Liu, D., Yu, L., Wei, L., Yu, P., Wang, J., Zhao, H., et al. (2021). BnTIR: an online transcriptome platform for exploring RNA-seq libraries for oil crop *Brassica napus*. *Plant Biotechnol. J.* 19, 1895–1897. doi: 10.1111/pbi.13665

Liu, H., Gu, F., Dong, S., Liu, W., Wang, H., Chen, Z., et al. (2016). *CONSTANS-like 9 (COL9)* delays the flowering time in *Oryza sativa* by repressing the Ehd1 pathway. *Biochem. Biophys. Res. Commun.* 479, 173–178. doi: 10.1016/j.bbrc.2016.09.013

Liu, J., Shen, J., Xu, Y., Li, X., Xiao, J., and Xiong, L. (2016). *Ghd2*, a *CONSTANS-like* gene, confers drought sensitivity through regulation of senescence in rice. *J. Exp. Bot.* 67, 5785–5798. doi: 10.1093/jxb/erw344

Lysak, M. A., Koch, M. A., Pecinka, A., and Schubert, I. (2005). Chromosome triplication found across the tribe Brassiceae. *Genome Res.* 15, 516–525. doi: 10.1101/gr.3531105

Min, J. H., Chung, J. S., Lee, K. H., and Kim, C. S. (2015). The *CONSTANS-like 4* transcription factor, *AtCOL4*, positively regulates abiotic stress tolerance through an abscisic acid-dependent manner in *Arabidopsis*. *J. Integr. Plant Biol.* 57, 313–324. doi: 10.1111/jipb.12246

Parkin, I. A., Lydiate, D. J., and Trick, M. (2002). Assessing the level of collinearity between *Arabidopsis thaliana* and *Brassica napus* for *A. thaliana* chromosome 5. *Genome* 45, 356–366. doi: 10.1139/g01-160

Reddy, B. A., Etkin, L. D., and Freemont, P. S. (1992). A novel zinc finger coiled-coil domain in a family of nuclear proteins. *Trends Biochem. Sci.* 17, 344–345. doi: 10.1016/0968-0004(92)90308-v

Robson, F., Costa, M. M., Hepworth, S. R., Vizir, I., Piñeiro, M., Reeves, P. H., et al. (2001). Functional importance of conserved domains in the flowering-time gene *CONSTANS* demonstrated by analysis of mutant alleles and transgenic plants. *Plant J.* 28, 619–631. doi: 10.1046/j.1365-313x.2001.01163.x

Roux, F., Touzet, P., Cuguen, J., and Le Corre, V. (2006). How to be early flowering: an evolutionary perspective. *Trends Plant Sci.* 11, 375–381. doi: 10.1016/j.tplants.2006.06.006

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425. doi: 10.1093/oxfordjournals.molbev.a040454

Sheng, P., Wu, F., Tan, J., Zhang, H., Ma, W., Chen, L., et al. (2016). A *CONSTANS-like* transcriptional activator, *OsCOL13*, functions as a negative regulator of flowering downstream of OsphyB and upstream of Ehd1 in rice. *Plant Mol. Biol.* 92, 209–222. doi: 10.1007/s11103-016-0506-3

Shim, J. S., Kubota, A., and Imaizumi, T. (2017). Circadian Clock and Photoperiodic Flowering in *Arabidopsis*: *CONSTANS* is a hub for signal Integration. *Plant Physiol.* 173, 5–15. doi: 10.1104/pp.16.01327

Song, J. M., Guan, Z., Hu, J., Guo, C., Zhang, Z., Wang, S., et al. (2020). Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nat. Plants* 6, 34–45. doi: 10.1038/s41477-019-0577-7

Song, N., Xu, Z., Wang, J., Qin, Q., Jiang, H., Si, W., et al. (2018). Genome-wide analysis of maize *CONSTANS-LIKE* gene family and expression profiling under light/dark and abscisic acid treatment. *Gene* 673, 1–11. doi: 10.1016/j.gene.2018.06.032

Steinbach, Y. (2019). The *Arabidopsis thaliana CONSTANS-LIKE 4* (*COL4*) - a modulator of flowering time. *Front. Plant Sci.* 10:651. doi: 10.3389/fpls.2019.00651

Takase, T., Kakikubo, Y., Nakasone, A., Nishiyama, Y., Yasuhara, M., Yoko, T. O., et al. (2012). Characterization and transgenic study of *CONSTANS-LIKE8* (*COL8*) gene in *Arabidopsis thaliana*: expression of 35S:COL8 delays flowering under long-day conditions. *Plant Tissue Cult. Lett.* 28, 439–446. doi: 10.5511/plantbiotechnology.11.0823b

Tan, J., Jin, M., Wang, J., Wu, F., Sheng, P., Cheng, Z., et al. (2016). *OsCOL10*, a *CONSTANS-Like* Gene, functions as a flowering time repressor downstream of Ghd7 in rice. *Plant Cell Physiol.* 57, 798–812. doi: 10.1093/pcp/pcw025

Tan, J., Wu, F., and Wan, J. (2017). Flowering time regulation by the CONSTANS-Like gene OsCOL10. *Plant Signal. Behav.* 12:e1267893. doi: 10.1080/15592324.2016.1267893

Tang, H., Wang, X., Bowers, J. E., Ming, R., Alam, M., and Paterson, A. H. (2008). Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res.* 18, 1944–1954. doi: 10.1101/gr.080978.108

Town, C. D., Cheung, F., Maiti, R., Crabtree, J., Haas, B. J., Wortman, J. R., et al. (2006). Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. *Plant Cell* 18, 1348–1359. doi: 10.1105/tpc.106.041665

Wang, H., Zhang, Z., Li, H., Zhao, X., Liu, X., Ortiz, M., et al. (2013). CONSTANS-LIKE 7 regulates branching and shade avoidance response in *Arabidopsis*. *J. Exp. Bot.* 64, 1017–1024. doi: 10.1093/jxb/ers376

Wenkel, S., Turck, F., Singer, K., Gissot, L., Le Gourrierec, J., Samach, A., et al. (2006). CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*. *Plant Cell* 18, 2971–2984. doi: 10.1105/tpc.106.043299

Wu, F., Price, B. W., Haider, W., Seufferheld, G., Nelson, R., and Hanzawa, Y. (2014). Functional and evolutionary characterization of the *CONSTANS* gene family in short-day photoperiodic flowering in soybean. *PLoS One* 9:e85754. doi: 10.1371/journal.pone.0085754

Wu, W., Zhang, Y., Zhang, M., Zhan, X., Shen, X., Yu, P., et al. (2018). The rice CONSTANS-like protein OsCOL15 suppresses flowering by promoting Ghd7 and repressing RID1. *Biochem. Biophys. Res. Commun.* 495, 1349–1355. doi: 10.1016/j.bbrc.2017.11.095

Wu, W., Zheng, X. M., Chen, D., Zhang, Y., Ma, W., Zhang, H., et al. (2017). *OsCOL16*, encoding a CONSTANS-like protein, represses flowering by up-regulating Ghd7 expression in rice. *Plant Sci.* 260, 60–69. doi: 10.1016/j.plantsci.2017.04.004

Xiao, G., Li, B., Chen, H., Chen, W., Wang, Z., Mao, B., et al. (2018). Overexpression of *PvCO1*, a bamboo *CONSTANS-LIKE* gene, delays flowering by reducing expression of the *FT* gene in transgenic *Arabidopsis*. *BMC Plant Biol.* 18:232. doi: 10.1186/s12870-018-1469-0

Yang, S., Weers, B. D., Morishige, D. T., and Mullet, J. E. (2014). *CONSTANS* is a photoperiod regulated activator of flowering in sorghum. *BMC Plant Biol.* 14:148. doi: 10.1186/1471-2229-14-148

Yang, T., He, Y., Niu, S., Yan, S., and Zhang, Y. (2020). Identification and characterization of the *CONSTANS* (CO)/*CONSTANS-like* (COL) genes related to photoperiodic signaling and flowering in tomato. *Plant Sci.* 301:110653. doi: 10.1016/j.plantsci.2020.110653

Zhang, Z., Ji, R., Li, H., Zhao, T., Liu, J., Lin, C., et al. (2014). *CONSTANS-LIKE 7* (*COL7*) is involved in phytochrome B (phyB)-mediated light-quality regulation of auxin homeostasis. *Mol. Plant* 7, 1429–1440. doi: 10.1093/mp/ssu058

Check for updates

Dynamic Surface Tension Enhances the Stability of Nanobubbles in Xylem Sap

Stephen Ingram[1]*, Yann Salmon[1,2], Anna Lintunen[1,2], Teemu Hölttä[2], Timo Vesala[1,2,3] and Hanna Vehkamäki[1]

[1] Institute for Atmospheric and Earth System Research/Physics, University of Helsinki, Helsinki, Finland, [2] Institute for Atmospheric and Earth System Research/Forest Sciences, University of Helsinki, Helsinki, Finland, [3] Laboratory of Ecosystem-Atmospheric Interactions of Forest – Mire Complexes, Yugra State University, Khanty-Mansiysk, Russia

Air seeded nanobubbles have recently been observed within tree sap under negative pressure. They are stabilized by an as yet unidentified process, although some embolize their vessels in extreme circumstances. Current literature suggests that a varying surface tension helps bubbles survive, but few direct measurements of this quantity have been made. Here, we present calculations of dynamic surface tension for two biologically relevant lipids using molecular dynamics simulations. We find that glycolipid monolayers resist expansion proportionally to the rate of expansion. Their surface tension increases with the tension applied, in a similar way to the viscosity of a non-Newtonian fluid. In contrast, a prototypical phospholipid was equally resistant to all applied tensions, suggesting that the fate of a given nanobubble is dependent on its surface composition. By incorporating our results into a Classical Nucleation Theory (CNT) framework, we predict nanobubble stability with respect to embolism. We find that the metastable radius of glycolipid coated nanobubbles is approximately 35 nm, and that embolism is in this case unlikely when the external pressure is *less negative* than −1.5 MPa.

Keywords: tree hydraulics, molecular simulation (molecular modeling), nanobubbles, lipid monolayers, Classical Nucleation Theory (CNT)

INTRODUCTION

Trees are capable of transporting water at high flow rates against gravity, without the aid of a mechanical pump like the heart in animals. Water vapor transpires from the leaves, generating tension (Pickard, 1981; Zimmermann et al., 1994), which pulls water further upwards in the xylem. This mechanism has come to be known as the cohesion-tension theory (Pockman et al., 1995). Severe tension or, more accurately, negative pressure, may disrupt the hydrogen bonds linking water molecules to one another, in much the same way that exposure to vacuum boils liquid water (Vera et al., 2016).

Nanobubbles further complicate the picture: it was recently discovered that gas-filled bubbles, tens to hundreds of nanometers in radius, can be "air seeded" into the liquid xylem sap at pit membranes (Schenk et al., 2015; Kaack et al., 2019). Intuitively, one would expect nanobubbles to be unstable with respect to "boiling" at highly negative pressures. Consider that a bubble's enthalpy,

H, can be decomposed into its internal energy, U, and the mechanical work required to expand the bubble volume, $v(r)$, at radius r, against the external pressure, p:

$$H(r) = U(r) + pv(r) \qquad (1)$$

We can see that when the external pressure is negative, the mechanical work of bubble formation is also negative. Therefore, the enthalpy of formation is decreased (Menzl and Dellago, 2016) relative to a positive pressure, and the growth of existing bubbles stabilizes the system energetically (see also **Supplementary Material**).

Classical Nucleation Theory (CNT) tells us that creating or maintaining a gas-liquid interface requires enthalpy, whereas increasing the bubble volume maximizes entropy. The critical radius of a bubble is the size that balances these two contributions: Surface contributions dominate at small radii, below the critical size, and volume contributions above. Any bubble exceeding the critical radius should undergo runaway growth, forcing dissolved gas from the liquid phase into the gas phase and embolizing the xylem conduit surrounding it. Yet this is not the case: Tree sap has been reported to be more stable under tension than water (Schenk et al., 2017; Yang et al., 2020).

Recent experiments (Ohgaki et al., 2010) on nitrogen nanobubbles, prepared under standard conditions, have revealed that they can sustain positive internal pressures of at least 6 MPa (60 atmospheres) and remain stable. Only very recently have theoretical explanations of this phenomena (Manning, 2020) been attempted, and it remains unclear how nanobubbles are prevented from expanding into embolisms. On a more fundamental level, perhaps it is pertinent to consider whether the presence of a small number of nanobubbles within xylem sap may actually be beneficial for trees.

To our knowledge, the only computational study of the effects of negative pressure on a biological system thus far has been that of Kanduč et al. (2020) Bilayers of amphiphilic lipids were subjected to very high tensions (−20 MPa) and were found to cavitate (nucleate pockets of vacuum) at rates predictable by CNT. In addition, Molecular Dynamics (MD) studies of pure water (Abascal et al., 2013; Gonzalez et al., 2015) have found that pressures more negative than –100 MPa are required to promote cavitation at an observable rate. Such conditions are too extreme to be relevant to tree sap: Cavities will not form spontaneously in xylem.

It has been suggested that nanobubbles will rapidly become coated in lipids as they "bud off" from the pit membrane structure (Schenk et al., 2017). Their presence is significant: As Schenk and co-workers have pointed out (Schenk et al., 2017; Yang et al., 2020), the dynamic surface tension of a surfactant coated gas-liquid interface, under tension, is unknown. Nonetheless, the structural changes undergone by monolayers as they are stretched at positive pressures are well established (Duncan and Larson, 2008). We predict that phase transitions within the nanobubble coating are common at the pressures found in trees. Furthermore, we hypothesize that their surface tension will change as different monolayer morphologies are adopted (Wüstneck et al., 2005).

In this study, we have used MD simulations to investigate the dynamical behavior of air-lipid-water interfaces as they are stretched at biologically relevant negative pressures. The interfaces are flat and smaller than the internal surface area of a prototypical nanobubble, creating a "toy" system free from the effects of curvature, or bubble dissolution. We have produced pressure—area isotherms, which express the dynamic nature of surface tension as a function of the instantaneous area per lipid. Finally, we directly relate this quantity to the Gibbs free energy landscape of nanobubbles, predicting the pressure range within which embolism is likely.

MATERIALS AND METHODS

Simulation Setup

All molecular dynamics simulations described herein were conducted using the 2020.2 iteration of GROMACS (Hess et al., 2008) running on the Puhti supercomputer at the CSC. GPU acceleration was provided using the CUDA platform (Kutzner et al., 2015), and each simulation utilized two NVIDIA Tesla v100 cards.
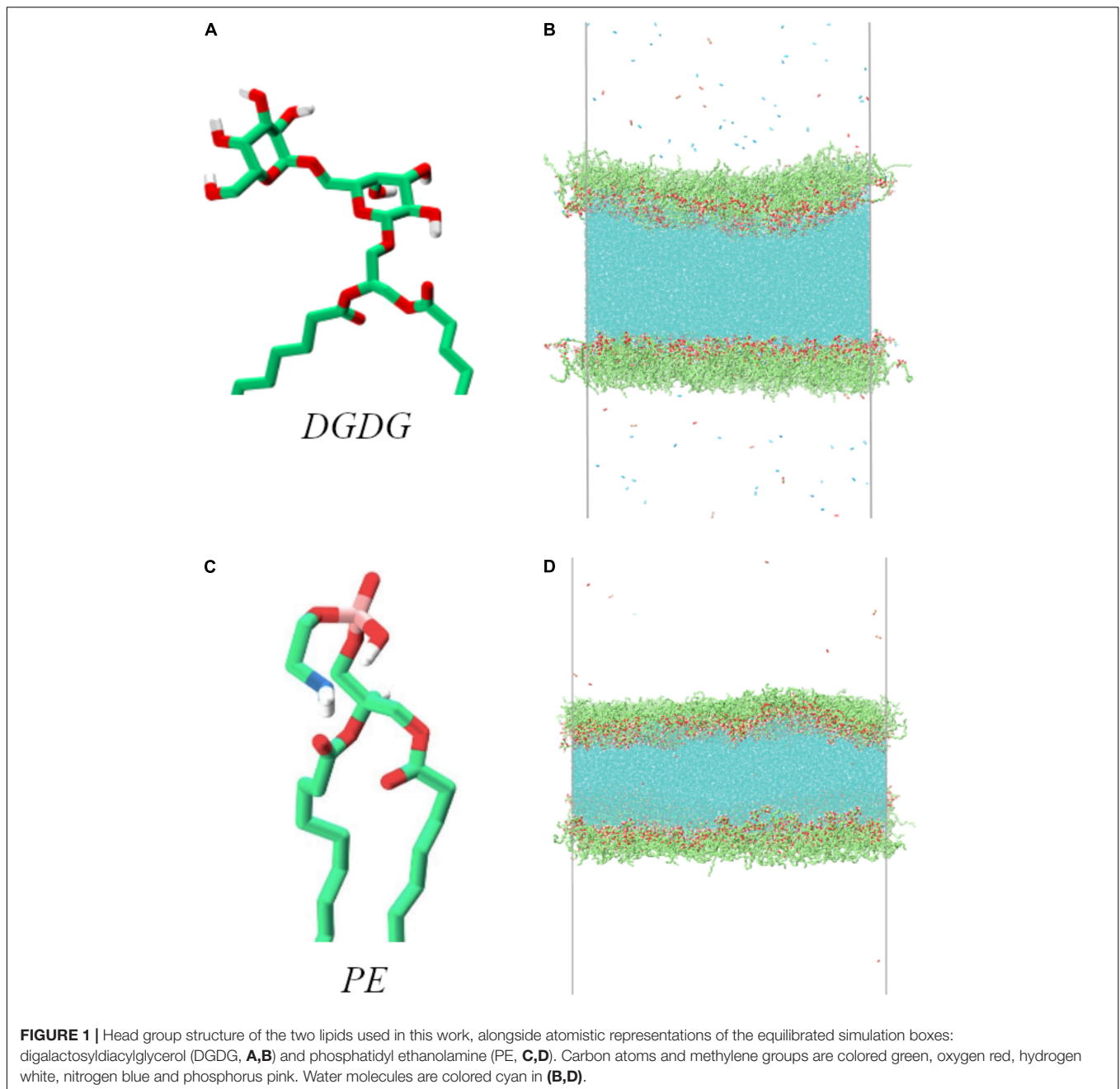
The simulations conducted consist of a horizontal slab of water coated with two monolayers of lipids decorating its top and bottom sides (in the z dimension). The lipids investigated were digalactosyldiacylglycerol (DGDG) and phosphatidyl ethanolamine (PE). The structure of their head groups can be seen in **Figure 1**. The two tail groups, extending below the bottom of the figure, are palmitic acid esters. The two gas phase volumes above and below the monolayers were filled with O_2 and CO_2 in a 50:50 ratio (for the DGDG) and N_2, O_2, and CO_2 in an 80:10:10 ratio (for PE), at pressures of 0.02 and 0.1 MPa, respectively (approximately 1 molecule per 50 nm^3). The difference in gas phase molecules included is not thought to affect the results at this pressure, as none of them were observed to reside to the water-lipid interface.

The major advantage of a double interface configuration is that any quantity we are interested in extracting (in this case surface tension and rupture area) can be determined for each monolayer separately, doubling the data extracted per simulation and reducing uncertainties in the desired quantities.

Our simulations were conducted in the NpT thermodynamic ensemble, which is to say that the number of atoms, their mean temperature, and the applied pressure are conserved over time. For both lipid systems, we have used a temperature of 298 K. We aim to understand the effect that different pulling rates have on the predicted nanobubble stability, beginning with the integrity of the interface itself.

The molecular configurations of the lipids were downloaded from the Limonada website.[1] To increase simulation efficiency, both lipids were included as united atom molecules: all hydrogen atoms present in the tail groups were omitted and their masses and charges folded into ("united" with) the appropriate carbon atoms.

---

[1] https://limonada.univ-reims.fr/

FIGURE 1 | Head group structure of the two lipids used in this work, alongside atomistic representations of the equilibrated simulation boxes: digalactosyldiacylglycerol (DGDG, **A,B**) and phosphatidyl ethanolamine (PE, **C,D**). Carbon atoms and methylene groups are colored green, oxygen red, hydrogen white, nitrogen blue and phosphorus pink. Water molecules are colored cyan in **(B,D)**.

Both lipids studied here were simulated using a variant of the GROMOS 53a6 force field, described by van Eerden et al. (2015). Water was represented by the TIP4P-2005 model, which effectively replicates the experimental properties of bulk water (Vega and de Miguel, 2007; Abascal and Vega, 2010). In the gas phase, CO_2 was represented by the EPM2 model (Cygan et al., 2012), and O_2 and N_2 by assigning the atoms the appropriate GROMOS atom type. The topology files used for both gas phase molecules were downloaded from the Automated Topology Builder[2] (Koziara et al., 2014).

With regards to pressure coupling, the Berenden barostat was used semi-isotropically, i.e., acting only in the lateral xy plane. The size of the box in the z dimension (thickness of the gas phase) is kept constant (60 nm for DGDG simulations, 45 nm for PE). The lateral compressibility of the all membranes investigated was fixed at a value of 5×10^{-7} MPa^{-1}, taken from the study of the lipid DPPC by Duncan and Larson (2008). The twin range cutoffs were set at 0.8 nm for DGDG simulations, while PE monolayers were simulated with both 0.8 and 1.6 to test whether a shorter value would properly represent the P–N interactions within the PE head group. The dynamic behavior of the monolayers was found to be identical in both cases.

[2]https://atb.uq.edu.au/

The original configurations were constructed in stages, using the packmol program (Martínez et al., 2009). In the DGDG simulations, the water slab contained 40,000 molecules in an approximate volume of $18 \times 18 \times 6$ nm. For PE, the number was 60,000. The two surfactant monolayers decorating the slab were then constructed by semi-random insertions of DGDG (276 molecules per monolayer) or PE (280 molecules per monolayer), with two restrictions: (1) That no two atoms be closer than 2 Å from one another, and (2) That the head group atoms were always placed closer to the water slab than the tail group atoms.

These configurations were energy minimized using the steepest descents algorithm, to allow the bonds and angles to relax closer to the constraints in the topology file. Next, the pressure and temperature coupling were activated, and the system equilibrated for 4 ns to generate the input coordinates for the data collection runs. Velocities were generated to fit a Maxwell Boltzmann distribution at 298 K, with the random seed changed in each simulation.

Pore Area Calculation

To understand the stability and integrity of the monolayers, a script was written to chart the exposed area of water, in two dimensions, as a function of time. Taking inspiration from the works of Gonzalez et al. (2015) and Wang and Frenkel (2005), the cavities in the monolayer were determined by comparison of the head group positions with a two-dimensional grid, as shown in **Figure 2**. Briefly, the algorithm works as follows:

Step one: An appropriate atom is chosen to represent the head group position of each lipid. In this case, phosphorus and nitrogen were chosen for PE, and all oxygens present within the galactosyl groups for DGDG. The x and y coordinates of these N atoms are then read in using the mdtraj python module (McGibbon et al., 2015).
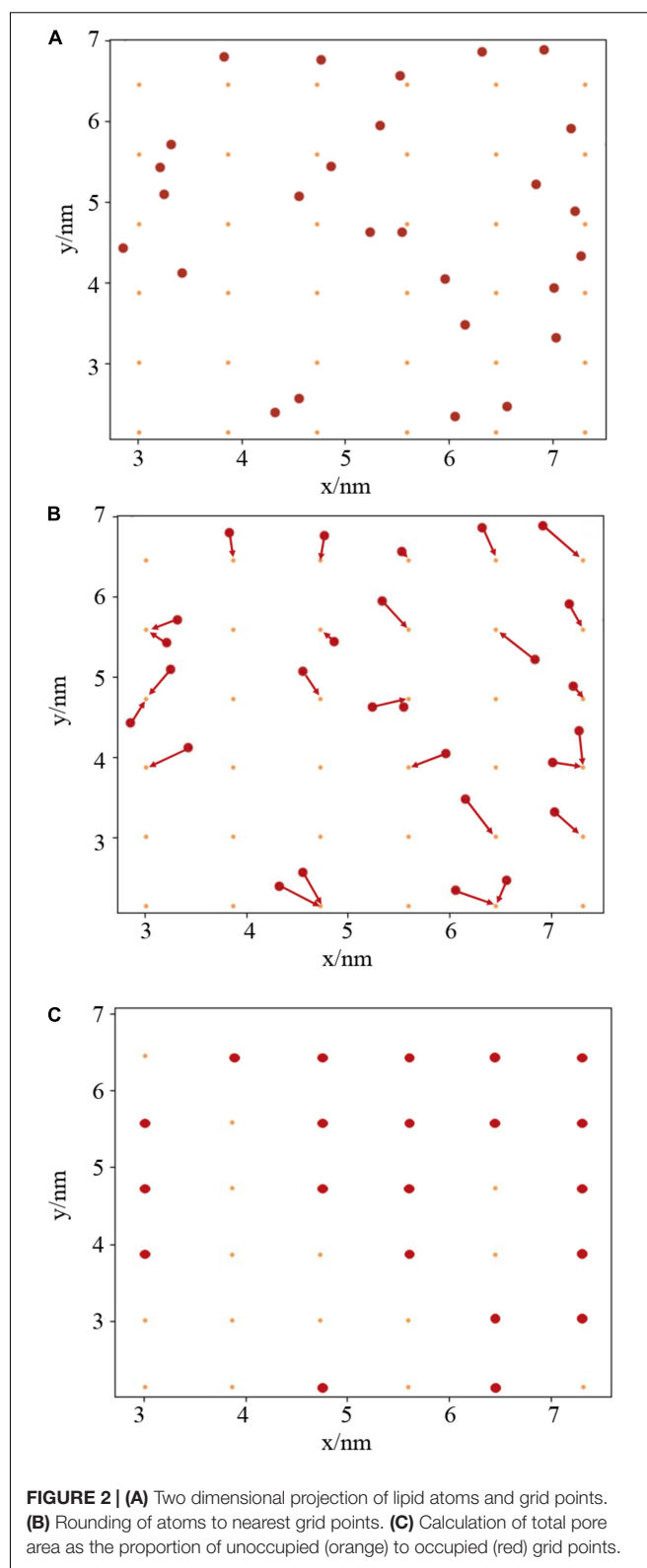
Step two: create an $M \times M$ grid, slightly offset from the box walls, where M^2 is approximately equal to the number of atoms read in above (panel **A**). Generate a corresponding $M \times M$ counting matrix, populated with zeros.

Step three: For each time step, round the atomic coordinates to the nearest grid point (panel **B**). This is achieved by dividing every point by the grid spacing, and then converting to an integer. If more than one atom is present at each grid point, add one to the corresponding element of the counting matrix. The pore area at each timestep is then calculated as proportion of zeros in the counting matrix (panel **C**).

The advantage of this algorithm over that of Gonzalez, or Voronoi tessellation methods like those used by Wang and Frenkel (2005) to study monolayer rupture, is that at no point are the grid positions directly compared with atom positions, nor are any magnitudes calculated. This grants significant speed boosts when analyzing long trajectories since both of those procedures scale much more strongly with M and N than division of a vector by a scalar.

Surface Tension Calculation

Unless otherwise stated, the value of γ_{water} used in this study was 40 mNm^{-1}, as opposed to the real value of 72 mNm^{-1}.

FIGURE 2 | (A) Two dimensional projection of lipid atoms and grid points. **(B)** Rounding of atoms to nearest grid points. **(C)** Calculation of total pore area as the proportion of unoccupied (orange) to occupied (red) grid points.

This is based on underpredictions of γ_{water} by the TIP4P-2005 water model, when simulated with twin cutoff values of 0.8 nm (Javanainen et al., 2018). Surface tension has been calculated in

two ways: The pressure tensor method for the main analysis, and the capillary wave method as a sanity check.

The pressure tensor (PT) method (Vega and de Miguel, 2007) involves subtracting the normal elements, P_x and P_y, from the perpendicular element, P_z, of the pressure tensor as a function of time.

$$\gamma_{PT} = \frac{1}{n} h_z \left[P_z - \frac{1}{2} \left(P_x + P_y \right) \right] \qquad (2)$$

where h_z is the vertical height of the simulation box, and n is the number of interfaces present (in this case 2). In simulations when the surface tension is expected to be constant, or to approach equilibrium as time proceeds, one can calculate γ_{PT} as a cumulative sum.

The capillary wave (CW) method relies on the intrinsic relationship between the surface tension and the width of the interface (Nickerson et al., 2013). It is useful here as it allows independent determination of the tensions of each monolayer separately. It requires the calculation of two length scales which have dubious physical meaning: the interfacial variance, or fluctuation, and the molecular diameter.

$$\gamma_{CW} = \frac{k_B T}{2\pi\sigma^2} \ln \left(\frac{h_z}{l} \right), \qquad (3)$$

where σ is the interfacial variance, and l is the molecular diameter. Here, we have calculated σ using the lateral density of the simulation box, by fitting an error function. Molecular diameter is calculated as the mean of the water Van der Waals diameter (2.8 Å) and the diameter of gyration of the lipid head groups,

$$l = 2 \sqrt{\frac{h_{xy}^2}{\pi n_{lipids}}} \qquad (4)$$

where h_{xy} is the simulation box width. Therefore, l varies as a function of time. γ_{CW} is calculated in 500 ps chunks, by dividing the z dimension of the simulation box into 300 slices.

A comparison of the time dependent values of γ_{CW} with the cumulative mean of γ_{PT} for one of the DGDG simulations conducted is shown in **Figure 3**. Given that σ in Equation (3) is a function of molecular coordinates, γ_{CW} begins much lower than γ_{PT}, and increases over the first few nanoseconds, proportionally with the pulling rate. By contrast, γ_{PT} is dependent only on the atomic forces, and so adopts larger values at earlier times.

The two methods converge, and a reasonable agreement is reached at around 4 ns. At no point in any simulation did the γ_{CW} estimates of the top and bottom monolayers deviate by more than the variance within one 500 ps bin, or show dissimilar trends. Therefore, the pressure tensor value was considered an accurate representation of the "mean" monolayer behavior, and so was chosen for the analysis.

RESULTS

In the current study, we applied four biologically relevant negative pressures to the glycolipid digalactosyldiacylglycerol (DGDG), between –0.5 and –3.5 MPa inclusive. DGDG was
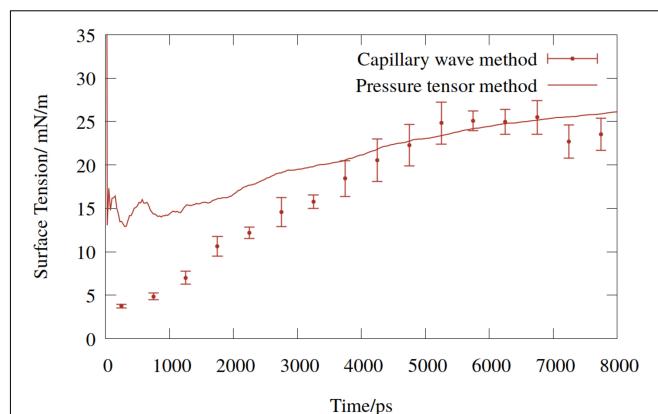

FIGURE 3 | Time dependent surface tension of digalactosyldiacylglycerol (DGDG) monolayers (mean of upper and lower) pulled at a pressure of −1.5 MPa. Solid line represents the cumulative average of the pressure tensor estimate, points represent the capillary wave method.
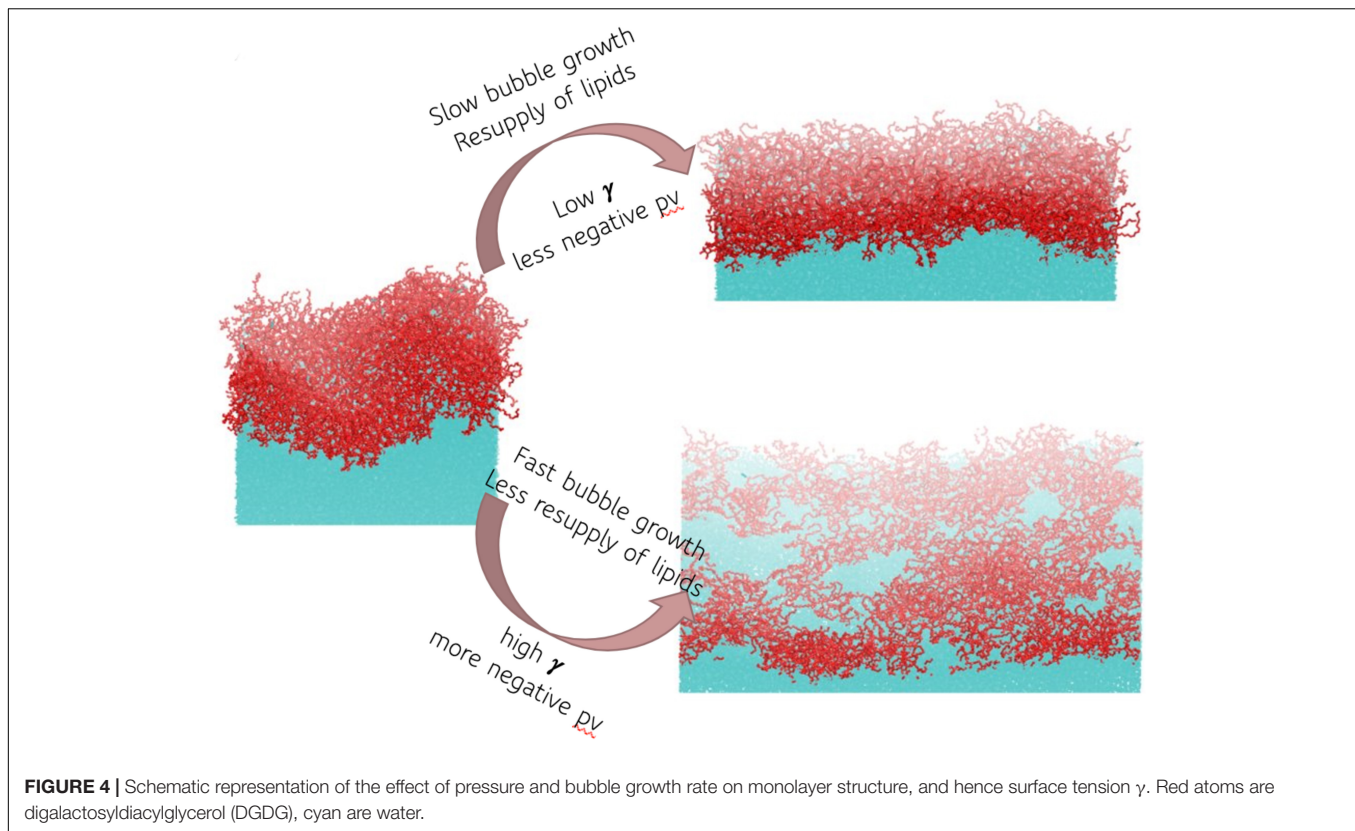
identified in the sap in a recent landmark publication (Schenk et al., 2021) using electrospray ionization mass spectrometry to characterize the lipid composition of xylem sap of seven angiosperm species. In all but the lowest tension case (−0.5 MPa) rupture of the monolayer was observed, exposing the water surface below. Snapshots of one DGDG interface before and after pulling at two example pressures is shown in **Figure 4**. Note the square nature of the interface, and the radically different areas of water exposed in the two right hand schematics.

The phospholipid Phosphatidyl ethanolamine (PE), another lipid present in sap (Schenk et al., 2021), was also investigated. As we will see in the following section, its monolayers were found to be completely resistant to pulling down to pressures of −5.5 MPa, which exceeds the range of tensions commonly found in plants.

Digalactosyldiacylglycerol Monolayer Rupture and Pore Formation

Both experimental and computational (Baoukina et al., 2007; Javanainen et al., 2018), studies have shown that lipid monolayers transition through several phases, depending on the area available: from liquid-condensed L_c, to liquid expanded L_e, to a ruptured network of pores surrounded by otherwise intact regions of L_e [for schematic representations of the phases we recommend the publications of Baoukina et al. (2007) and Duncan and Larson (2008)]. Note that unlike previous literature, the monolayers here are being actively pulled to mimic negative xylem pressures, rather than simply allowed to relax to a reduced *positive* pressure.

A script was written to determine the area of exposed water in every frame of the MD trajectory (as described in section "Materials and Methods"). This analysis was applied to the lipid DGDG and the results are presented in **Figure 5A**. We can see that a more negative pressure causes a faster increase in the pore area within the lipid monolayer during the pulling. The monolayers experiencing the most modest pressure (−0.5 MPa) maintained a sufficient lateral force that the water

FIGURE 4 | Schematic representation of the effect of pressure and bubble growth rate on monolayer structure, and hence surface tension γ. Red atoms are digalactosyldiacylglycerol (DGDG), cyan are water.

slab was stopped from embolizing (boiling). Interestingly, at the three pressures during which rupture was observed, the nanobubble monolayers lost integrity at the same pore coverage: approximately 10% of the surface area. Once that threshold was reached, runaway growth took hold, signifying embolism would have occurred at that point within a xylem vessel.

Dynamic Surface Tension and Surface Pressure

In the initial configuration of the DGDG simulations, the monolayer is highly doped and buckles, or warps, slightly to fit all of the lipids in. Such a configuration is a consequence of the fact that the head group area is significantly larger than the cross-sectional area of the tail (Israelachvili, 2011). Concomitantly, the starting surface tension, at 5.3 mNm^{-1}, is extremely low (see **Figure 3**). By contrast, PE is more capable of lamellar packing at this length scale, due to the head group being closer in area to the footprint of the tail.

In the literature, surface *pressure*, Π, is the preferred metric for quantifying interfacial properties. We have chosen to express our results in this format. Conversion of surface tension to surface pressure was achieved by subtraction of the calculated surface tension from the water surface tension determined by the same method, for the same model:
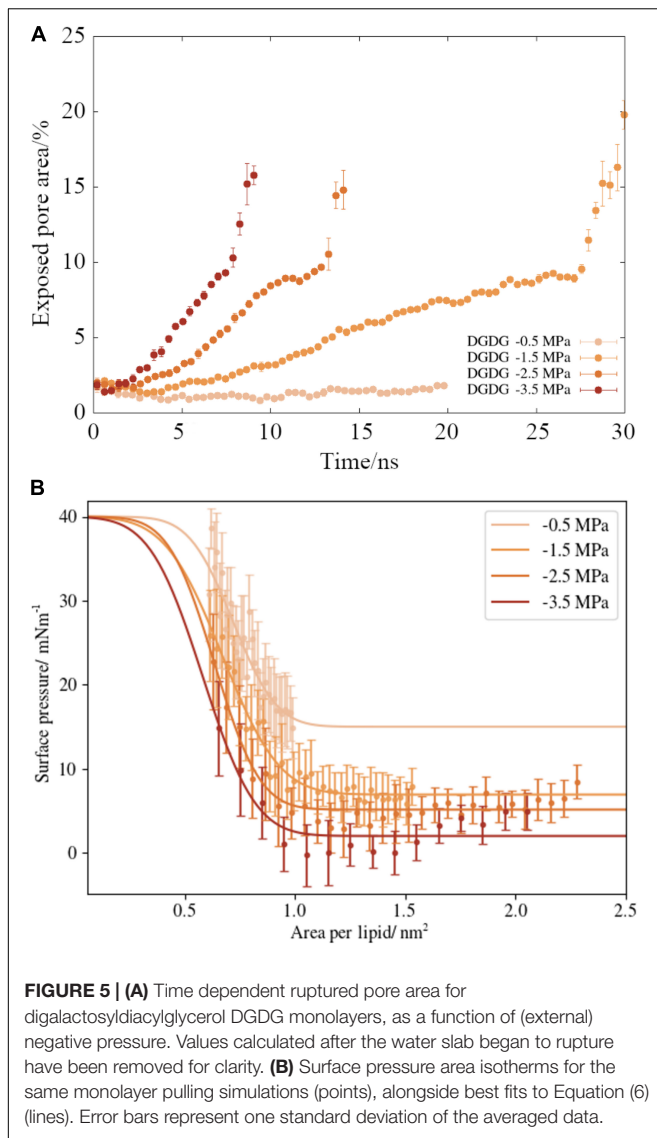
$$\Pi(A) = \gamma_{water} - \gamma_{interface}(A) \qquad (5)$$

Upon application of negative pressure coupling, the interface is stretched at rates proportional to the applied tension. Therefore, the ability of the lipids to maintain order is disrupted and, as shown in **Figures 4**, **5A**, the monolayers lose structural integrity. We present the surface pressure area isotherms (named by analogy with pressure volume isotherms of gases) of DGDG in **Figure 5B**, which shows that Π (*A*) begins to decrease toward zero (the point at which the interface behaves as if no lipids were present) rapidly in all four simulations. Indeed, for the −3.5 MPa pressure, the observed surface pressure reaches zero in the region 0.9–1.5 lipids nm^{-2}.

Time averaged quantities, such as Π, fluctuate significantly during molecular dynamics simulations, which can be a problem when extrapolating the data to predict properties of larger systems. In order to present the CNT free energy as a smooth function of radius in section "Discussion," the calculated pressure area isotherms should be parameterized as simple, monotonically varying functions. Ideally, a physically realistic representation of Π(*A*) will plateau at low *A*, where the monolayer is stable. It will then transition to a much lower value at high *A*, as the monolayer ruptures. The error function was identified as containing these features, and was thus selected for this purpose:

$$\Pi(A) = \Pi_0 - \frac{\Pi_{rupture}}{2} \cdot (1 + \mathrm{erf}(b(A - c))) \qquad (6)$$

Here Π_0 is the surface pressure before the transition, $\Pi_{rupture}$ is the surface pressure *difference* between the ruptured and

FIGURE 5 | (A) Time dependent ruptured pore area for digalactosyldiacylglycerol DGDG monolayers, as a function of (external) negative pressure. Values calculated after the water slab began to rupture have been removed for clarity. **(B)** Surface pressure area isotherms for the same monolayer pulling simulations (points), alongside best fits to Equation (6) (lines). Error bars represent one standard deviation of the averaged data.

unruptured configurations. During the fitting of Equation (6) to the data, the fit parameter b and c were introduced to determine the shape and center of the transition, respectively. We chose to fix the value of Π_0 to be γ_{water}—ε, where ε is the mean absolute error of the surface pressure for each external pressure (i.e., the mean size of the bars in **Figure 5B**). The fitting was carried out using the curve_fit function in python, with the following bounds: $0.1 < b < 5$, $0.5 < c < 1.2 \; nm^{-2}$, and $2 < \Pi_{rupture} < 20 \; mNm^{-1}$.

The best-fit curves are presented alongside the appropriate data points in **Figure 5B**. The influence of external pressure on surface pressure can be seen by the fact that $\Pi_{rupture}$, the width of the transition, increases monotonically as p becomes more negative.

Calculating the corrected $\gamma_{interface}$ (A) values using Equation (5) produces a "true" surface tension range accessed by the DGDG simulations of between 36 and 73 mNm^{-1} inclusive of all pressures. We note that the range of corrected $\gamma_{interface}$ values

exhibited at -1.5 MPa pressure is 46–67 mNm^{-1}, which is almost identical to the bulk xylem sap surface tensions experimentally measured by Losso et al. (2017) from the gynosperm species *Picea abies* and *Pinus mugo*. Bulk xylem measurements in angiosperms (Christensen-Dalsgaard et al., 2011) have produced slightly larger values (55–70 mNm^{-1}), although it should be noted that bulk sap will have a lower average lipid concentration than a nanobubbles internal surface.

There is a small increase in the calculated surface pressure at $0.8 \; nm^{-2}$ in the -0.5 MPa case (**Figure 5B**). Similar "activation barriers" have been observed in MD simulations of monolayers before (Javanainen et al., 2018), and are indicative of the transition from a mostly L_c phase structure to a mostly L_e. The presence of such a barrier in our data gives confidence we have accurately modeled the phase characteristics of the system. It also suggests that monolayers experiencing less negative pressures behave more conventionally than those under more extreme tensions.

Stability of Phosphatidyl Ethanolamine Monolayers

Three monolayer pulling simulations were conducted using the phospholipid PE, each beginning from the coordinates shown in **Figure 1D**, at external pressures -1.5, -3.5, and -5.5 MPa. We found PE monolayers to be highly resistant to pulling: Even tensions exceeding those commonly observed in xylem (-5.5 MPa) (Lintunen et al., 2013; Losso et al., 2017) were not able to produce significant changes in monolayer area on the timescale of the simulation (the lateral box size increased by less than a 2% over 22.5 ns). Similarly, while the calculated surface tensions, γ_{PT}, were different in each simulation, the values fluctuated much less, as a function of time, relative to the glycolipid monolayers.

By exhibiting both lateral stability and numerically stable surface tensions, it follows that PE monolayers occupy a single point in surface pressure-area space, rather than the sigmoidal function we used to represent the glycolipid DGDG in **Figure 5B**. We therefore consider the behaviors of these two lipids to be different in kind, rather than degree (at least within the pressure range studied). The implications of this for mixed monolayers will be discussed in the next section. For the time being, we can say that the surfactant (or surfactants) present on the surface of a nanobubble will radically alter its fate within the xylem.

The two-dimensional structure of the PE monolayer is also distinct from DGDG: The head groups of PE do not organize themselves across the interface at a uniform density. Instead, they adopt a two-dimensional structure that resembles filaments when viewed from above. Much of the surface water is not bonded directly to a head group, leading to an equilibrium pore area of $\sim$23%. The benefit of such a configuration appears to be to maximize the contact between the P and P/N atoms of adjacent head molecules, which may account for the lateral stability of PE.

Comparison of the two lipid head groups shows that they have a similar number of neighbors, as determined by calculating their coordination numbers (2.35 for PE vs. 2.59 for DGDG). However, the first solvation shell of PE is significantly smaller than that

of DGDG (6 vs. 15 Å), implying that the phospholipid bonds more closely and strongly to itself. We show a snapshot of the configuration, alongside the relevant radial distribution functions of Phorphorus and Nitrogen in **Supplementary Figure 5**.

DISCUSSION

While this study and our results belong to computational or soft matter physics, they have important implications for nanobubble stability under tension and thus, for plant physiology. Therefore, we would like to expand and elaborate on some of the connections.

Pressure-Area Isotherms

Returning to **Figure 5B**, we can see that the four sets of points do not overlap significantly in pressure/area space. Instead, more negative external pressures lead to lower surface pressure at a given bubble size. We can therefore infer that the lipid structure has less time to re-organize in response to the interface expanding at more negative pressures, as the area increases faster than the lateral diffusion of lipids can compensate for.

What these results show is that there is a negative feedback loop present at the surface: The more rapid the expansion of the monolayer, the larger the increase in surface tension. Paradoxically, more negative pressures will lead to more stable bubble surfaces with respect to embolism, because the energetic cost of continuing to grow the surface is much higher. One can make an analogy here to a non-Newtonian fluid: a slow deformation will be met with minimal resistance, whereas the application of a great force will cause an instant vitrification in kind. We therefore propose that this effect is the crucial factor determining the fate of each bubble. By extension, if trees are able to regulate lipid concentration such that a certain bubble size is favored, there may be the means by which trees have evolved to tolerate some bubbles entering the xylem, while suppressing their ability to embolize.

Classical Nucleation Theory

We have just shown that the surface of an expanding nanobubble will act in such a way as to counteract that expansion. However, the Gibbs free energy of a nucleation process is also a function of the volume of the new phase. Is the surface tension increase sufficient to counteract the mechanical work extracted, pv, by increasing the gas volume per bubble? To answer this, we have calculated the formation free energy of a hypothetical DGDG covered nanobubble using a CNT approach:

$$g(r) = 4\pi r^2 \frac{\gamma_{interface}(r)}{1 + 2\delta/r} + pv(r) \qquad (7)$$

Here r is the bubble radius, v its volume and δ the Tolman length. p is the pressure difference between the internal bubble pressure and the external negative pressure. The size dependent surface tensions, $\gamma_{interface}(r)$, were produced from the error function parameterizations presented in section " Materials and Methods." To convert γ from a function of area per lipid into a

function of bubble radius we have assumed that, at each pressure, the bubbles will have a unique number of lipids, n_{lipids}, coating their inner surface. Here we have chosen $n_{lipids} = 30,000$ for all four pressures. We present in the **Supplementary Information** two further sets of $g(r)$ curves, assuming n_{lipids} is pressure dependent. We find that the general shape of the curves remains similar.

A note on the Tolman length: δ is an abstract model parameter with no concrete experimental counterpart, and a wide range of values have been proposed in studies modeling nanobubbles, or water under negative pressure: They range from –0.047 nm (Azouzi et al., 2013; Yasui et al., 2016) to 17 nm (Manning, 2020). Here we have chosen 0.2 nm, the value used by Menzl et al. (2016) to simulate the cavitation of pure water.

The formation free energy curves calculated for DGDG coated bubbles at four negative pressures are shown in **Figure 6**. All exhibit a two barrier "kinetic trap" type dependence on bubble radius [visible as local minima in $g(r)$]. The first potential barrier, which can be seen in the insert figure, is the barrier to homogeneous cavitation, i.e., formation of a small volume of vacuum with a surface tension close to zero (Abascal et al., 2013). The second barrier is the barrier to embolism. In all cases, the metastable radius is approximately 35 nm: a bubble of this size lies at a local minimum of Gibbs free energy where it is thermodynamically unstable relative to a fully embolized vessel, but kinetically trapped from achieving embolism. A separate calculation of surface entropy across this radius range is presented in **Supplementary Material**.

Metastability in nanobubbles is not a novel concept: It was proposed recently by Yarom and Marmur (2015), who hypothesized that a bubble of insoluble gas can become kinetically stabilized when its size is equal to the critical radius. Such a mechanism is similar to that experienced by small
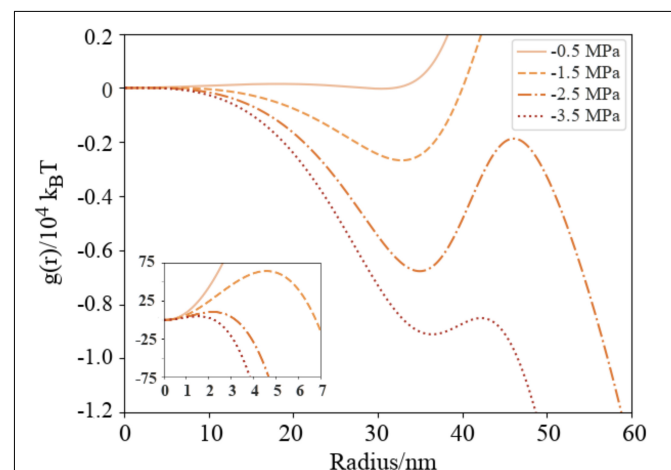

FIGURE 6 | Naïve Gibbs free energy surfaces (Equation 7) for DGDG as a function of external (negative) pressures in a nanobubble with varying radius, presented in units of 10^4 k_BT. $P_{external}$ = -0.5 MPa (solid line), −1.5 MPa (dashed), −2.5 (dash-dotted) and −3.5 (dotted). Insert figure shows cavitation free energy barriers in the region $r = 0$–7 nm in units of $k_B T$. Color scheme is the same as **Figure 5**.

aerosol particles, and described by the so-called Kohler theory (Köhler, 1936). By contrast, we are proposing here a nanobubble containing soluble gas achieving metastability at a potential energy minimum, rather than maximum.

A literature survey was conducted to ascertain whether a size distribution centered at 35 nm was physically realistic given the air seeding model of bubble formation. Our results were found to compare very favorably with Boutilier et al. (2014), who filtered solid particulates through a pine branch, extracting a distribution centered at a radius of 40 nm. By contrast, Choat et al. (2003) observed that only nanoparticles between 5 and 20 nm were able to pass through a pit membrane. This may not preclude the existence of larger bubbles, however, as they could adopt an elongated shape during seeding, with a small cross-sectional area.

The −0.5 and −1.5 MPa curves contain enormous barriers to embolism, meaning that were a distribution of air-seeded bubbles to be subject to either energy landscape, they would *never embolize*. Instead, each bubble would either expand or contract from its original size until it reached the local minimum. Any further attempt to expand would be met with significant resistance, even if lipids could be resourced from nearby to stave off monolayer rupture.

Conversely, any reduction in radius from the local minimum also destabilizes the bubbles: they approach the activation barrier for cavitation (the peaks at ∼5 nm in the inset **Figure 6**) from the other side. Physically, one would expect a concomitant increase in internal pressure to occur, and hence an outward force acting to restore the radius to its metastable value, reducing the likelihood of sub 10 nm bubbles. The existence of these barriers support the conclusions of the recent publication by Vehmas and Makkonen (2021), which suggests that bubble dissolution becomes unlikely at small radii as the internal pressures increase significantly.

As Schenk et al. (2017) have pointed out, surface area to volume ratios dictate that fragmentation into several smaller bubbles is energetically favored over expansion, assuming curvature effects can be discounted. We therefore propose that, at xylem sap tensions between 0 and at least –1.5 MPa, the dynamic behavior of nanobubbles is characterized by a mechanism of continuous expansion, fragmentation, and recycling into smaller bubbles. Mechanistically, what occurs may resemble the transitions between nanobubbles and liposomes observed by Koshiyama and Wada (2016), where the coating monolayers buckle and fold inward, creating multiple internal surfaces inside the bubble. Clearly, further work is needed to more fully validate the existence of this process on a microscopic level.

For the −2.5 and −3.5 MPa cases, the peaks of the barriers to embolism are below the 0 free energy level. Physically speaking, this means that the critical bubble will be more stable than the same amount of gas dissolved within the liquid. Therefore, bubbles should have sufficient thermal energy to embolize at these pressures, assuming the energy is not immediately dissipated into the surrounding tissue, or tapped by the tree in some other way.

A similar phenomenon has been observed in bimolecular chemical reactions (Glowacki et al., 2011), whereby products retain vibrational or thermal energy for extended periods, without dissipation, allowing them to undergo further transformations. Alternatively, if the residual Gibbs free energy is quickly drained, the bubbles will adopt the local minimum radius and remain metastable, like those experiencing lower tensions. We therefore consider the rate of energy dissipation directly after air seeding occurs to be a key parameter in determining the stability of nanobubbles in the xylem vessels.

Given that tension within the xylem sap increases from the roots to the leaves, one could propose that a bubble moving upwards within a xylem vessel will be subject to each of these energy surfaces in turn. In such a situation, the mechanical work of formation becomes more negative, but this is offset by the interfacial tension of the bubble increasing. Furthermore, nanobubbles will cross multiple pit membranes on their journey upwards, which could change their coating.

An increase in the surface tension of a nanobubble, but no increase in size, may allow for gradually higher internal gas pressures to be sustained, potentially as high as those observed by Ohgaki et al. (2010). The Laplace pressure, defined as

$$p_{Laplace} = \frac{2\gamma(r)}{r} \tag{8}$$

allows us to estimate the equilibrium vapor pressure within a spherical bubble. A few instructive examples: Inside a 35 nm bubble with a $\gamma(r)$ of 47 mNm^{-1} ($\Pi = 25$ mNm^{-1}), $P_{Laplace}$ is +2.7 MPa, rising to +4.1 MPa for a $\gamma(r)$ of 72 mNm^{-1} (a plot of estimated $P_{Laplace}$ against $P_{external}$ is provided in **Supplementary Figure 3**). These values suggest that nanobubbles under high tension can transport more gas by molar quantity than those under less negative external pressures, even if they are the same size. We note that the critical pressure of N_2 is 3.4 MPa (Lemmon et al., 2021), meaning that in extreme cases the air present in nanobubbles may in fact condense into a supercritical state. Additionally, CO_2 and O_2, which are both present in xylem sap, are believed to be moderately surface active at high pressures, and so should migrate to the interface whenever any pores in the monolayer form. The precise interplay of these phenomena, and their effects on bubble stability, remains elusive.

Possible Effects of Mixed Monolayers

It is highly likely that the surface coating of each nanobubble within xylem sap contains multiple different surfactants, at varying concentrations, or even small proteins such as saponins (Christensen-Dalsgaard et al., 2011) which are known to interact with lipid bilayers in plant cell walls (Lin and Wang, 2010). Indeed, given that the lipids present at pit membranes mostly originate from cells that have died, it stands to reason that their phase behavior echoes, to an extent, that of the membranes from which they came.

Combining lipids which are highly resistant to pulling with those more easily ruptured could potentially allow the organism to tune the pressure range (for which there is a huge variability from species to species; Choat et al., 2012) within which bubbles can survive. Our results hint at such a variability: Of the two

lipids investigated herein, only one (DGDG) accommodated any expansion of its monolayers at the pressures investigated. By contrast, PE was shown to have remarkable resilience within the range of biologically relevant tensions (we note that PE was a minor constituent, relative to DGDG, in all seven angiosperms investigated by Schenk et al., 2021). The behavior of other phospholipids under tension, or indeed mixed monolayers of glyco- and phospholipids, remains elusive.

CONCLUSION

Here we have used state of the art molecular dynamics methods to directly calculate the surface tension of two biologically relevant lipids that can be expected to coat nanobubbles at a range of negative pressures found in trees. The values calculated for the glycolipid DGDG are in line with previous measurements of bulk xylem sap surface tension. We discover that lipid monolayers are less capable of rearranging at more negative pressures, destabilizing the internal surface but stabilizing the nanobubble with respect to runaway growth. CNT predicts that, at external pressures between 0 and –1.5 MPa, this effect is sufficient to avoid embolism altogether. At more negative pressures, embolism becomes increasingly likely, depending on the rate of free energy dissipation. We propose a mechanism wherein xylem nanobubbles repeatedly expand and collapse into smaller bubbles, recycling their surface lipids in the process. These novel results reconcile the existence of nanobubbles in xylem sap with the cohesion-tension theory of water transport in plants.

To our knowledge, this is the first time a dynamic surface tension has been shown to be influenced by pulling rate (i.e., varying external tensions) rather than by more common hysteresis effects which depend on whether the interface is expanding or compressing. We are also unaware of any direct comparisons of surface energy expended to the mechanical work gained by boiling in the context of air seeding, rather than homogeneous cavitation. Further work will investigate the impact of significantly elevated Laplace pressures within the bubble on interfacial tensions. In the future, we aim to explore the phenomenon of lipid resupply to an expanding interface, in the form of micelles or other self-assembled structures, and the rate at which it proceeds.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

SI conducted the simulations and analysis, under the supervision of HV and TV. SI wrote the majority of the manuscript. YS and TH were part of the same academy of Finland funding consortium, and wrote small parts of the manuscript during revisions. Along with AL, they shaped the project through multiple interdisciplinary meetings throughout 2019 and 2020. All authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.732701/full#supplementary-material

REFERENCES

Abascal, J. L. F., and Vega, C. (2010). Widom line and the liquid-liquid critical point for the TIP4P/2005 water model. *J. Chem. Phys.* 133:234502. doi: 10.1063/1.3506860

Abascal, J. L. F., Gonzalez, M. A., Aragones, J. L., and Valeriani, C. (2013). Homogeneous bubble nucleation in water at negative pressure: a voronoi polyhedra analysis. *J. Chem. Phys.* 138:084508. doi: 10.1063/1.4790797

Azouzi, M. E. M., Ramboz, C., Lenain, J.-F., and Caupin, F. A. (2013). Coherent picture of water at extreme negative pressure. *Nat. Phys.* 9, 38–41.

Baoukina, S., Monticelli, L., Marrink, S. J., and Tieleman, D. P. (2007). Pressure-area isotherm of a lipid monolayer from molecular dynamics simulations. *Langmuir* 23, 12617–12623. doi: 10.1021/la702286h

Boutilier, M. S. H., Lee, J., Chambers, V., Venkatesh, V., and Karnik, R. (2014). Water filtration using plant xylem. *PLoS One* 9:e89934. doi: 10.1371/journal.pone.0089934

Choat, B., Ball, M., Luly, J., and Holtum, J. (2003). Pit membrane porosity and water stress-induced cavitation in four co-existing dry rainforest tree species. *Plant Physiol.* 131, 41–48. doi: 10.1104/pp.014100

Choat, B., Jansen, S., Brodribb, T. J., Cochard, H., Delzon, S., Bhaskar, R., et al. (2012). Global convergence in the vulnerability of forests to drought. *Nature* 491, 752–755. doi: 10.1038/nature11688

Christensen-Dalsgaard, K. K., Tyree, M. T., and Mussone, P. G. (2011). Surface tension phenomena in the xylem sap of three diffuse porous temperate tree species. *Tree Physiol.* 31, 361–368. doi: 10.1093/treephys/tpr018

Cygan, R. T., Romanov, V. N., and Myshakin, E. M. (2012). Molecular simulation of carbon dioxide capture by montmorillonite using an accurate and flexible force field. *J. Phys. Chem. C* 116, 13079–13091. doi: 10.1021/jp3007574

Duncan, S. L., and Larson, R. G. (2008). Comparing experimental and simulated pressure-area isotherms for DPPC. *Biophys. J.* 94, 2965–2986. doi: 10.1529/biophysj.107.114215

Glowacki, D. R., Rose, R. A., Greaves, S. J., Orr-Ewing, A. J., and Harvey, J. N. (2011). Ultrafast energy flow in the wake of solution-phase bimolecular reactions. *Nat. Chem.* 3, 850–855. doi: 10.1038/nchem.1154

Gonzalez, M. A., Abascal, J. L. F., Valeriani, C., and Bresme, F. (2015). Bubble nucleation in simple and molecular liquids via the largest spherical cavity method. *J. Chem. Phys.* 142:154903. doi: 10.1063/1.4916919

Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008). GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4, 435–447. doi: 10.1021/ct700301q

Israelachvili, J. N. (ed). (2011) "Soft and Biological Structures," in *Intermolecular and Surface Forces*, ed. J. N. Israelachvili (Amsterdam: Elsevier), 535–576. doi: 10.1016/B978-0-12-375182-9.10020-X

Javanainen, M., Lamberg, A., Cwiklik, L., Vattulainen, I., and Ollila, O. H. S. (2018). Atomistic model for nearly quantitative simulations of langmuir monolayers. *Langmuir* 34, 2565–2572. doi: 10.1021/acs.langmuir.7b02855

Kaack, L., Altaner, C. M., Carmesin, C., Diaz, A., Holler, M., Kranz, C., et al. (2019). Function and three-dimensional structure of intervessel pit membranes in angiosperms: a review. *IAWA J.* 40, 673–702. doi: 10.3390/plants9020231

Kanduč, M., Schneck, E., Loche, P., Jansen, S., Schenk, H. J., and Netz, R. R. (2020). Cavitation in lipid bilayers poses strict negative pressure stability limit in biological liquids. *Proc. Natl. Acad. Sci.* 117, 10733–10739. doi: 10.1073/pnas.1917195117

Köhler, H. (1936). The nucleus in and the growth of hygroscopic droplets. *Trans. Faraday Soc.* 32, 1152–1161.

Koshiyama, K., and Wada, S. (2016). Collapse of a lipid-coated nanobubble and subsequent liposome formation. *Sci. Rep.* 6:28164. doi: 10.1038/srep28164

Koziara, K. B., Stroet, M., Malde, A. K., and Mark, A. E. (2014). Testing and validation of the automated topology builder (ATB) version 2.0: prediction of hydration free enthalpies. *J. Comput. Aided. Mol. Des.* 28, 221–233. doi: 10.1007/s10822-014-9713-7

Kutzner, C., Páll, S., Fechner, M., Esztermann, A., de Groot, B. L., and Grubmüller, H. (2015). Best bang for your buck: GPU nodes for GROMACS biomolecular simulations. *J. Comput. Chem.* 36, 1990–2008. doi: 10.1002/jcc.24030

Lemmon, E. W., McLinden, M. O., and Friend, D. G. (2021). *Thermophysical Properties of Fluid Systems. In NIST Chemistry WebBook, NIST Standard Reference Database Number 69*. Gaithersburg, MD: National Institute of Standards and Technology.

Lin, F., and Wang, R. (2010). Hemolytic mechanism of dioscin proposed by molecular dynamics simulations. *J. Mol. Model.* 16, 107–118. doi: 10.1007/s00894-009-0523-0

Lintunen, A., Hölttä, T., and Kulmala, M. (2013). Anatomical regulation of ice nucleation and cavitation helps trees to survive freezing and drought stress. *Sci. Rep.* 3, 1–7. doi: 10.1038/srep02031

Losso, A., Beikircher, B., Dämon, B., Kikuta, S., Schmid, P., and Mayr, S. (2017). Xylem sap surface tension may be crucial for hydraulic safety. *Plant Physiol.* 175, 1135–1143. doi: 10.1104/pp.17.01053

Manning, G. S. (2020). On the thermodynamic stability of bubbles, immiscible droplets, and cavities. *Phys. Chem. Chem. Phys.* 22, 17523–17531. doi: 10.1039/d0cp02517h

Martínez, L., Andrade, R., Birgin, E. G., and Martínez, J. M. (2009). PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* 30, 2157–2164. doi: 10.1002/jcc.21224

McGibbon, R. T., Beauchamp, K. A., Harrigan, M. P., Klein, C., Swails, J. M., Hernández, C. X., et al. (2015). MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* 109, 1528–1532. doi: 10.1016/j.bpj.2015.08.015

Menzl, G., and Dellago, C. (2016). Effect of entropy on the nucleation of cavitation bubbles in water under tension. *J. Chem. Phys.* 145:211918. doi: 10.1063/1.4964327

Menzl, G., Gonzalez, M. A., Geiger, P., Caupin, F., Abascal, J. L. F., Valeriani, C., et al. (2016). Molecular mechanism for cavitation in water under tension. *Proc. Natl. Acad. Sci. U.S.A.* 113, 13582–13587. doi: 10.1073/pnas.1608421113

Nickerson, S., Frost, D. S., Phelan, H., and Dai, L. L. (2013). Comparison of the capillary wave method and pressure tensor route for calculation of interfacial tension in molecular dynamics simulations. *J. Comput. Chem.* 34, 2707–2715. doi: 10.1002/jcc.23443

Ohgaki, K., Khanh, N. Q., Joden, Y., Tsuji, A., and Nakagawa, T. (2010). Physicochemical approach to nanobubble solutions. *Chem. Eng. Sci.* 65, 1296–1300. doi: 10.1016/j.ces.2009.10.003

Pickard, W. F. (1981). The ascent of sap in plants. *Prog. Biophys. Mol. Biol.* 37, 181–229.

Pockman, W. T., Sperry, J. S., and O'Leary, J. W. (1995). Sustained and significant negative water pressure in xylem. *Nature* 378, 715–716. doi: 10.1093/jxb/erm281

Schenk, H. J., Espino, S., Romo, D. M., Nima, N., Do, A. Y. T., Michaud, J. M., et al. (2017). Xylem surfactants introduce a new element to the cohesion-tension theory. *Plant Physiol.* 173, 1177–1196. doi: 10.1104/pp.16.01039

Schenk, H. J., Michaud, J. M., Mocko, K., Espino, S., Melendres, T., Roth, M. R., et al. (2021). Lipids in xylem sap of woody plants across the angiosperm phylogeny. *Plant J.* 105, 1477–1494. doi: 10.1111/tpj.15125

Schenk, H. J., Steppe, K., and Jansen, S. (2015). Nanobubbles: a new paradigm for air-seeding in xylem. *Trends Plant Sci.* 20, 199–205. doi: 10.1016/j.tplants.2015.01.008

van Eerden, F. J., de Jong, D. H., de Vries, A. H., Wassenaar, T. A., and Marrink, S. J. (2015). Characterization of thylakoid lipid membranes from cyanobacteria and higher plants by molecular dynamics simulations. *Biochim. Biophys. Acta Biomembr.* 1848, 1319–1330. doi: 10.1016/j.bbamem.2015.02.025

Vega, C., and de Miguel, E. (2007). Surface tension of the most popular models of water by using the test-area simulation method. *J. Chem. Phys.* 126:154707. doi: 10.1063/1.2715577

Vehmas, T., and Makkonen, L. (2021). Metastable nanobubbles. *ACS Omega* 6, 8021–8027. doi: 10.1021/acsomega.0c05384

Vera, F., Rivera, R., Romero-Maltrana, D., and Villanueva, J. (2016). Negative pressures and the first water siphon taller than 10.33 meters. *PLoS One* 11:e0153055. doi: 10.1371/journal.pone.0153055

Wang, Z.-J., and Frenkel, D. (2005). Pore nucleation in mechanically stretched bilayer membranes. *J. Chem. Phys.* 123:154701. doi: 10.1063/1.2060666

Wüstneck, R., Perez-Gil, J., Wüstneck, N., Cruz, A., Fainerman, V. B., and Pison, U. (2005). Interfacial properties of pulmonary surfactant layers. *Adv. Colloid Int. Sci.* 117, 33–58. doi: 10.1016/j.cis.2005.05.001

Yang, J., Michaud, J., Jansen, S., Schenk, H. J., and Zuo, Y. Y. (2020). Dynamic surface tension of xylem sap lipids. *Tree Physiol.* 40, 433–444. doi: 10.1093/treephys/tpaa006

Yarom, M., and Marmur, A. (2015). Stabilization of boiling nuclei by insoluble gas: can a nanobubble cloud exist? *Langmuir* 31, 7792–7798. doi: 10.1021/acs.langmuir.5b00715

Yasui, K., Tuziuti, T., and Kanematsu, W. (2016). Extreme conditions in a dissolving air nanobubble. *Phys. Rev.* 94, 1–13. doi: 10.1103/PhysRevE.94.013106

Zimmermann, U., Zhu, J. J., Meinzer, F. C., Goldstein, G., Schneider, H., Zimmermann, G., et al. (1994). High molecular weight organic compounds in the xylem sap of mangroves: implications for long-distance water transport. *Bot. Acta* 107, 218–229. doi: 10.1111/j.1438-8677.1994.tb00789.x

Check for updates

A Comprehensive Analysis of Calmodulin-Like Proteins of *Glycine max* Indicates Their Role in Calcium Signaling and Plant Defense Against Insect Attack

Manisha Yadav[1], Jyotsna Pandey[1], Amrita Chakraborty[2], Md. Imtaiyaz Hassan[3], Jiban Kumar Kundu[4], Amit Roy[2]*, Indrakant Kumar Singh[5,6]* and Archana Singh[1]*

[1] Department of Botany, Hansraj College, University of Delhi, New Delhi, India, [2] EVA4.0 Unit, Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Prague, Czechia, [3] Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi, India, [4] Plant Virus and Vector Interactions Group, Crop Research Institute, Prague, Czechia, [5] Molecular Biology Research Laboratory, Department of Zoology, Deshbandhu College, University of Delhi, New Delhi, India, [6] DBC-i4 Center, Deshbandhu College, University of Delhi, New Delhi, India

The calcium (Ca^{2+}) signaling is a crucial event during plant-herbivore interaction, which involves a transient change in cytosolic Ca^{2+} concentration, which is sensed by Ca^{2+}-sensors, and the received message is transduced to downstream target proteins leading to appropriate defense response. Calmodulin-like proteins (*CMLs*) are calcium-sensing plant-specific proteins. Although *CMLs* have been identified in a few plants, they remained uncharacterized in leguminous crop plants. Therefore, a wide-range analysis of *CMLs* of soybean was performed, which identified 41 true CMLs with greater than 50% similarity with *Arabidopsis CMLs*. The phylogenetic study revealed their evolutionary relatedness with known *CMLs*. Further, the identification of conserved motifs, gene structure analysis, and identification of *cis*-acting elements strongly supported their identity as members of this family and their involvement in stress responses. Only a few *Glycine max CMLs (GmCMLs)* exhibited differential expression in different tissue types, and rest of them had minimal expression. Additionally, differential expression patterns of *GmCMLs* were observed during *Spodoptera litura*-feeding, wounding, and signaling compound treatments, indicating their role in plant defense. The three-dimensional structure prediction, identification of interactive domains, and docking with Ca^{2+} ions of *S. litura*-inducible GmCMLs, indicated their identity as calcium sensors. This study on the characterization of *GmCMLs* provided insights into their roles in calcium signaling and plant defense during herbivory.

Keywords: Calmodulin like proteins (CMLs), soybean, *Spodoptera litura*, calcium signaling, signaling compounds, wounding, plant-insect interaction

INTRODUCTION

Plants being sessile are constantly attacked by various environmental cues that can cause abiotic or biotic stresses (Stotz et al., 2000). Among all, insect attack is crucial biotic stress, affecting the plants' fitness, performance, and productivity (Stotz et al., 2000; Conrath et al., 2002; Dodd et al., 2010; Hancock et al., 2015; Singh et al., 2018, 2020a,b, 2021c; Chen and Mao, 2020;

Keshan et al., 2021; Singh and Singh, 2021). To protect themselves, plants activate their defense mechanism and produce different defensive compounds like secondary metabolites, volatile organic compounds, protease inhibitors, and other antiherbivore chemicals (Gatehouse, 2002; Halitschke and Baldwin, 2004; Kliebenstein, 2004; Schuler, 2011; Vadassery et al., 2012b; Lortzing and Steppuhn, 2016; Singh et al., 2021a,c). This activation of biosynthesis of defensive compounds is governed by certain early events that facilitate the detection of herbivory and signal transduction. However, stimulus perception and signal transduction during herbivory is not understood well. Moreover, plants must discriminate among various environmental stimuli to trigger a specific response by activating downstream cellular signaling. Therefore, it is mandatory to determine the perception molecules, crucial players of signal transduction, and molecular mechanism behind plant defense against herbivory (Kong et al., 2013; Singh et al., 2016, 2020c; Kumar et al., 2020).

Plants recognize herbivore attack by wounding patterns and interaction of herbivore/damage-associated molecular patterns (HAMPs/DAMPs) with pattern recognition receptors (PRRs) (Mithöfer and Boland, 2008). This interaction activates appropriate downstream signal transduction pathways for achieving stimulus-specific responses. The process of coupling the perception of feeding herbivores to plant adaptive response involves transient changes in the intracellular Ca^{2+} concentration (Maffei et al., 2004; Thor and Peiter, 2014). During the process, an increase in Ca^{2+} levels occurs due to the transport of Ca^{2+} across the membrane or from subcellular organelles, which is facilitated by plasma membrane Ca^{2+}-permeable channels, such as cyclic nucleotide-gated channels (CNGCs), glutamate receptor-type cation channel (GLRs), stretch-activated Ca^{2+} channel (OSCAs), and MID1-complementing activity (MCA) families and vacuolar channel TWO-PORE CHANNEL1 (TPC1) (Maffei et al., 2004; Dodd et al., 2010; Vadassery et al., 2012a; Kiep et al., 2015; DeFalco et al., 2016; Vincent et al., 2017; Meena et al., 2019; DeFalco and Zipfel, 2021). This transport of Ca^{2+} causes spatial and temporal variations in cellular distribution, frequency, amplitude, kinetics of Ca^{2+}, and intracellular Ca^{2+} level leading to the generation of calcium signature, which encrypts information from primary stimuli to encode specific intracellular responses (Ranty et al., 2016). These calcium signals are decoded further by various calcium sensor proteins that consist of canonical Ca^{2+} binding EF-hand motif with conserved helix-loop-helix structure coordinating one Ca^{2+} (Lecourieux et al., 2002; Gifford et al., 2007). These proteins perceive and interpret Ca^{2+} signals through binding to Ca^{2+} ions leading to a change in their conformation and functions (Aldon et al., 2018). In plants, Ca^{2+} binding proteins are represented by complex families. They can be either Ca^{2+} responders, which convey the signal *via* enzymatic reactions, such as Ca^{2+}-dependent protein kinases (CDPKs), or the non-catalytic sensor relay proteins, such as Calmodulins (CaMs), Calmodulin-like proteins (CMLs), and calcineurin B-like proteins (CBLs), which are activated upon binding to Ca^{2+}, which promote downstream signaling (DeFalco et al., 2010).

Calmodulin like proteins are principal representative of Ca^{2+} sensor proteins (DeFalco et al., 2010; Kumar et al., 2020)

and are characterized by two to six EF-hand motifs and are known to evolve from ancestor CaM, sharing at least 15% amino acid identity (Perochon et al., 2011; Bender and Snedden, 2013). So far, CMLs are pointed out to be involved in the developmental process, plant immunity, and stress responses by targeting many kinases, ion transporters, metabolic enzymes, phosphatases, and transcription factors (Reddy et al., 2011). The EF-hand motifs show a high affinity toward cooperative binding with Ca^{2+} that expose the hydrophobic surface, which allows them to interact with downstream target proteins (Vadassery et al., 2012a). *CML*8, *CML*37, *CML*38, and *CML*39 are responsive to drought, salinity, and hormonal treatment in *Arabidopsis* (Vanderbeld and Snedden, 2007; Park et al., 2010). *CML*9 is regulated by abscisic acid (ABA) and *Pseudomonas syringae* infection (Magnan et al., 2008; Leba et al., 2012). *CML*42 is known to cause aberration in trichomes (Dobney et al., 2009), and *CML*24 is involved in alterations in flowering time and ion stress (Delk et al., 2005; Hubbard et al., 2008). Plants possess a repertoire of *CMLs*, which are unique to them with 50 members in *Arabidopsis* (McCormack et al., 2005), 52 in *Solanum lycopersicum* (Munir et al., 2016), 19 in *Lotus japonicas* (Liao et al., 2017), and 62 in *Vitis vinifera* (Vandelle et al., 2018). Few *CMLs* of *Arabidopsis thaliana* (*CML9, 11, 12, 16, 17, 23,* and *42*) are upregulated upon employing oral secretion of *Spodoptera littoralis* (Vadassery et al., 2012a,b). *CML*37 and *CML*42 act antagonistically to regulate stress responses in *Arabidopsis* by altering phytohormone signals (Heyer et al., 2021). However, the involvement of these Ca^{2+} sensors has not been investigated much in crop plants. To achieve more insights into the function of *CMLs* and recognize their involvement in plant defense against herbivory, it would be crucial to investigate them during soybean-*Spodoptera litura* interaction.

Glycine max, commonly known as soybean, is a dietary staple crop of Asian countries and a native legume plant of Southeastern Asia (Soyastats, 2010; Badole and Bodhankar, 2012; Badole and Mahamuni, 2013; Naresh et al., 2019). It consists of a balanced proportion of amino acids essential for the growth of the human body (Naresh et al., 2019). The genome of soybean is fully sequenced, and the processes of gene regulation, disease resistance, and nodulation are well defined (Gresshoff and Ferguson, 2017). However, insect pests can adversely affect the yield and quality of soybean (Cui et al., 1997). *S. litura* (common cutworm), a polyphagous insect, is one of the major destructive and widespread soybean pests throughout the Asia-Pacific region. Plants have employed various constitutive and induced defense strategies along with sophisticated signaling networks to combat predation by herbivores (Singh et al., 2021b). The induced responses are more effective for plants and lead to increased tolerance against herbivores (Stotz et al., 2000; Zhang and Yang, 2000; Fan et al., 2012). *S. litura*, a nocturnal moth and a serious polyphagous defoliator, infests soybean plants through the vigorous eating pattern of larvae (Singh et al., 2021b). The proteolytic activity of the gut serine protease of the larvae is mainly responsible for the significant damage to crops (Vasudev and Sohal, 2016). Therefore, it is essential to elucidate how *S. litura*-infestation induces defense signaling in *G. max* evoking the plant defense mechanism.

In the present study, a comprehensive investigation of *CMLs* of *G. max* was executed across the genome. We studied the genomic organization, evolutionary and phylogenetic relationships, gene structure analysis, and identification of EF-hand motifs of *Glycine max CMLs (GmCMLs)*. We also evaluated the mRNA levels of *GmCMLs* in different tissue types. The investigation on change in mRNA levels of *GmCMLs* during *S. litura*-infestation revealed the involvement of 36 *GmCMLs* in plant defense against herbivory. We also screened the promoter sequences of *GmCMLs* to identify *cis*-acting elements that revealed the presence of Jasmonic acid (JA)-response elements and salicylic acid (SA)-related elements. Further, transcript profiling of *S. litura*-inducible *CMLs* upon defense-signaling compound treatments was also executed, which revealed their differential expression. We also predicted the three-dimensional structure of *S. litura*-inducible GmCMLs and its interaction with Ca^{2+}, which confirmed their identity and involvement in Ca^{2+} signaling. This analysis has offered promising candidates that can function as an essential component of Ca^{2+} signaling during *G. max-S. litura* interaction and can be valuable for deciphering the early events involved in the upregulation of plant defense in soybean and beyond.

MATERIALS AND METHODS

Sequence Retrieval, Phylogenetic Analysis, and Similarity Search of Calmodulin-Like Proteins of *Glycine max*

The protein and gene sequences of *CMLs* of *G. max* were identified and retrieved from the soybean database[1] and phytozome database[2]. The DNA and peptide sequences of GmCMLs were evaluated manually utilizing BLASTN and BLASTP[3]. The base pair lengths and number of amino acids for each *GmCMLs* were retrieved from NCBI[4].

To study the evolutionary relationships among model plants, monocot, and leguminous dicot, amino acid sequences of homologous GmCMLs from *Medicago truncatula*, *Arabidopsis thaliana*, and *Oryza sativa* were retrieved from Phytozome (see text footnote 2), TAIR[5], and Oryzabase[6], respectively. The amino acid sequences were then submitted to multiple sequence alignment using Clustal Omega[7] (Madeira et al., 2019). The generated multiple sequence alignment was subjected to "construct" phylogenetic tree. The details of constructed phylogenetic tree were downloaded using "Download" phylogenetic data and saved in newick format. For better presentation of the phylogenetic tree, the saved phylogenetic tree was uploaded on iTOL[8] (Letunic and Bork, 2021).

---

[1] https://www.soybase.org/

[2] https://phytozome.jgi.doe.gov/pz/portal.html

[3] https://blast.ncbi.nlm.nih.gov/Blast.cgi

[4] https://www.ncbi.nlm.nih.gov

[5] https://www.arabidopsis.org/

[6] https://shigen.nig.ac.jp/rice/oryzabase/

[7] https://www.ebi.ac.uk/Tools/msa/clustalo/

[8] https://itol.embl.de/itol.cgi

The identity of GmCML peptide sequences was compared with the protein sequence of CMLs of *Arabidopsis thaliana* using the BLAST engine of the TAIR database[9].

Expression Analysis of *Glycine max CMLs* During *Spodoptera litura*-Infestation

Plant Growth

The soybean seeds (Pusa 9712) were washed, soaked for 2 h, and kept at 26−27°C in a dark chamber between 2 layers of tissue paper, with continuous moistening with distilled water for germination. After 3–4 days, the germinated seedlings were transferred to the sterilized soilrite containing pots. The seedlings were grown in a plant growth chamber under controlled conditions (photoperiod: 16 h light and 8 h dark; temperature: 26−27°C; humidity: 55–60%) and watered regularly.

Rearing and Maintenance of *Spodoptera litura*

The larvae of *S. litura* were procured from ICAR – National Bureau Of Agricultural Insect Resources, Bangalore, India, and reared in the laboratory using standard protocols at 27°C and 65–70% relative humidity on a 14/10 h light/dark cycle (Singh et al., 2021b). The freshly molted fourth instar larvae were starved for 12 h before releasing them on plants for experimentation and bioassays.

Spodoptera litura-Infestation and Mechanical Wounding on Soybean Plants and Sample Collection

For *S. litura*-infestation, one-month-old soybean plants were exposed to one larva per plant. Plants without any treatments were served as control. Mechanical wounding was performed according to Singh et al. (2008). The entire shoot of treated (sample) and untreated (control) plants, after 1 h of infestation/wounding from three independent experiments, was harvested and immediately snap-frozen at −80°C for further analysis.

RNA Isolation and cDNA Synthesis

Trizol (TRI) reagent (Sigma Aldrich, Spruce Street, St. Louis, MO, United States) was used to isolate RNA as per the given protocol (Singh et al., 2018, 2021b; Keshan et al., 2021). The RNA concentrations were measured on a spectrophotometer, and RNA integrity was checked by agarose gel electrophoresis. The isolated RNA was used to synthesize first-strand cDNA using an iScript cDNA synthesis kit (Bio-Rad, Hercules, CA, United States).

Primer Design and Gene Expression Analysis Using qPCR

Primer 3 (v.0.4.0)[10] tool was used to design real-time (RT) PCR primers. Applied Biosystems™ 7500 RT PCR System (Thermo Fisher Scientific, Waltham, MA, United States) was used to perform RT-qPCR. Each 10 µl reaction mix contained 1 µl of cDNA samples, 0.5 µl of 10 µM gene-specific primers, and 5 µl of SYBR green supermix (Bio-Rad, Hercules, CA,

---

[9] https://www.arabidopsis.org/Blast/

[10] http://bioinfo.ut.ee/primer3-0.4.0/

United States). The soybean elongation factor gene was used as an internal control. The qPCR data were analyzed utilizing the $2^{-\Delta\Delta CT}$ method. The $2^{-\Delta\Delta CT}$ values obtained by qPCR were transformed to $\log^2$ to lessen noise level. For statistical analysis, ANOVA and Tukey's test ($P < 0.05$) were applied and a criterion of greater than two-fold induction/reduction level was taken into consideration to select differentially expressing genes according to a previous report (Singh et al., 2008). The heat map was generated by Graphpad Prism Software[11] using qPCR data.

Expression Analysis of *Glycine max CMLs* Upon Treatment With Signaling Compounds

For treatment with signaling compounds, an equal volume of 100 μM of methyl-jasmonate, 50 μM ethephon, 5 μM salicylic acid, and only water/0.5% (v/v) ethanol in water (control) were sprayed on healthy soybean plants according to the previous studies (Stotz et al., 2000; Singh et al., 2008). The plants were maintained under the same conditions but in individual enclosures. The leaves of treated (sample) and untreated (control) plants were harvested after 1 h of treatment in biological triplicates and immediately snap-frozen at −80°C for further analysis. For gene expression analysis, qPCR for 36 *GmCMLs* (differentially expressed on *S. litura*-infestation) was performed, and the data were analyzed using the $2^{-\Delta\Delta CT}$ method. For statistical analysis, ANOVA and Tukey's test ($P < 0.05$) were applied and a criterion of greater than two-fold induction/reduction level was taken into consideration to select differentially expressing genes. The heat map was generated by Graphpad Prism Software (see text footnote 11) using qPCR data.

In silico Analysis of True *Glycine max CMLs*

The subcellular localization of different GmCMLs was predicted using WoLF PSORT[12]. The molecular weight and isoelectric point of each GmCMLs were computed using the "Compute pI/MW tool" of ExPASy[13]. The presence of EF-hands was checked using the InterPro database[14]. The MEME software[15] was used to detect and generate logo plots of the EF-hands motifs.

Exon-Intron Determination

The organization of exons and introns in *GmCMLs* was deciphered using Gene Structure Display server GSDS2.0[16] (Hu et al., 2015).

Tissue-Specific Gene Expression Profiling

The expression profile values [Reads/Kb/Million (RPKM) normalized data] for different tissues, that is, flower, young leaf, green pods, root, nodule, stem, and seed, were retrieved from SoyBase Expression Explorer[17] (Waese et al., 2017). The heat map was illustrated using RPKM values for tissue-specific expression profiling of *GmCMLs* on Graphpad Prism Software (see text footnote 11).

Identification of microRNA (miRNA) Targets of *Glycine max CMLs*

Plant small-RNA target (psRNATarget) analysis server was used to predict miRNA target sites in *GmCMLs* transcripts with default parameters. *G. max*-specific mature miRNA sequences were downloaded from the miRNA database (miRBase)[18] (Bhatia et al., 2019).

Interaction of *Glycine max* CMLs With Other Proteins

To deduce the direct and indirect interaction of GmCMLs with other proteins, the sequences of all GmCMLs were subjected to the String database[19] (Szklarczyk et al., 2021). The GmCML protein sequences were submitted as multiple proteins with organisms selected as *G. max*. The submission of sequences redirected to a new window, where proteins matching with input sequences along with identity, bitscore, and *e*-value were listed. Protein interaction list was saved by clicking on "Mapping", and further to find interaction networks proceeded with "continue."

Detection of Regulatory Elements in the Promoter Region of *Glycine max CMLs*

To detect the regulatory elements present in the 5′ upstream region of *GmCMLs*, 1 kb of the genomic DNA sequences before the initiation codon (ATG) of each *GmCMLs* were saved from NCBI and subjected to Plant Care analysis[20] (Lescot et al., 2002).

Three-Dimensional Structure Prediction of Selected *Glycine max* CMLs by Homology Modeling

The homology modeling of *S. litura*-inducible GmCMLs was conducted to predict their three-dimensional structure. The selected *S. litura*-inducible GmCMLs were used as query sequences for homology modeling and searched against the Protein Data Bank (PDB) to identify the closely related known protein sequence. The data were submitted to the Phyre[2] Protein Homology/Analogy recognition Engine http://www.sbg.bio.ic.ac.uk/$\sim$phyre2/html/page.cgi?id=index to construct the protein 3-D structure by homology modeling under intensive modeling mode (Kelley et al., 2015). Further, to check the stereochemical quality of protein tertiary structures, the PROCHECK server was used http://saves.mbi.ucla.edu/ (Laskowski et al., 1996).

---

[11]https://www.graphpad.com/scientific-software/prism

[12]https://www.genscript.com/wolf-psort.html

[13]https://web.expasy.org/compute_pi/

[14]https://www.ebi.ac.uk/interpro/

[15]http://meme.sdsc.edu/meme/meme.html

[16]http://gsds.gao-lab.org/

[17]https://www.soybase.org/expression/

[18]http://plantgrn.noble.org/psRNATarget/

[19]https://string-db.org/cgi/input?sessionId=bcguyAaLcQA5&input_page_show_search=on

[20]http://bioinformatics.psb.ugent.be/webtools/plantcare

Interaction of Selected *Glycine max* CMLs With Ca²⁺

To show the interaction between Ca²⁺ and GmCMLs, the InChI format of Ca²⁺ was derived from PubChem[21] (Kim et al., 2019). For docking studies, the InChI format was converted to.pdb using OpenBabel–Chemical file format converter[22] (O'Boyle et al., 2011). The interaction studies were performed for true GmCMLs that were differentially regulated by *S. litura*-infestation using PATCHDOCK https://bioinfo3d.cs.tau.ac.il/PatchDock/php.php (Schneidman-Duhovny et al., 2005). Different parameters used for docking studies were clustering root-mean-square deviation (RMSD) = 4, and the complex type is protein and small ligand. The top ten Patchdock docking results were further refined using FireDock[23] (Mashiach et al., 2008). The GmCMLs were docked with Ca²⁺ based on the number of EF-hand motifs. The global
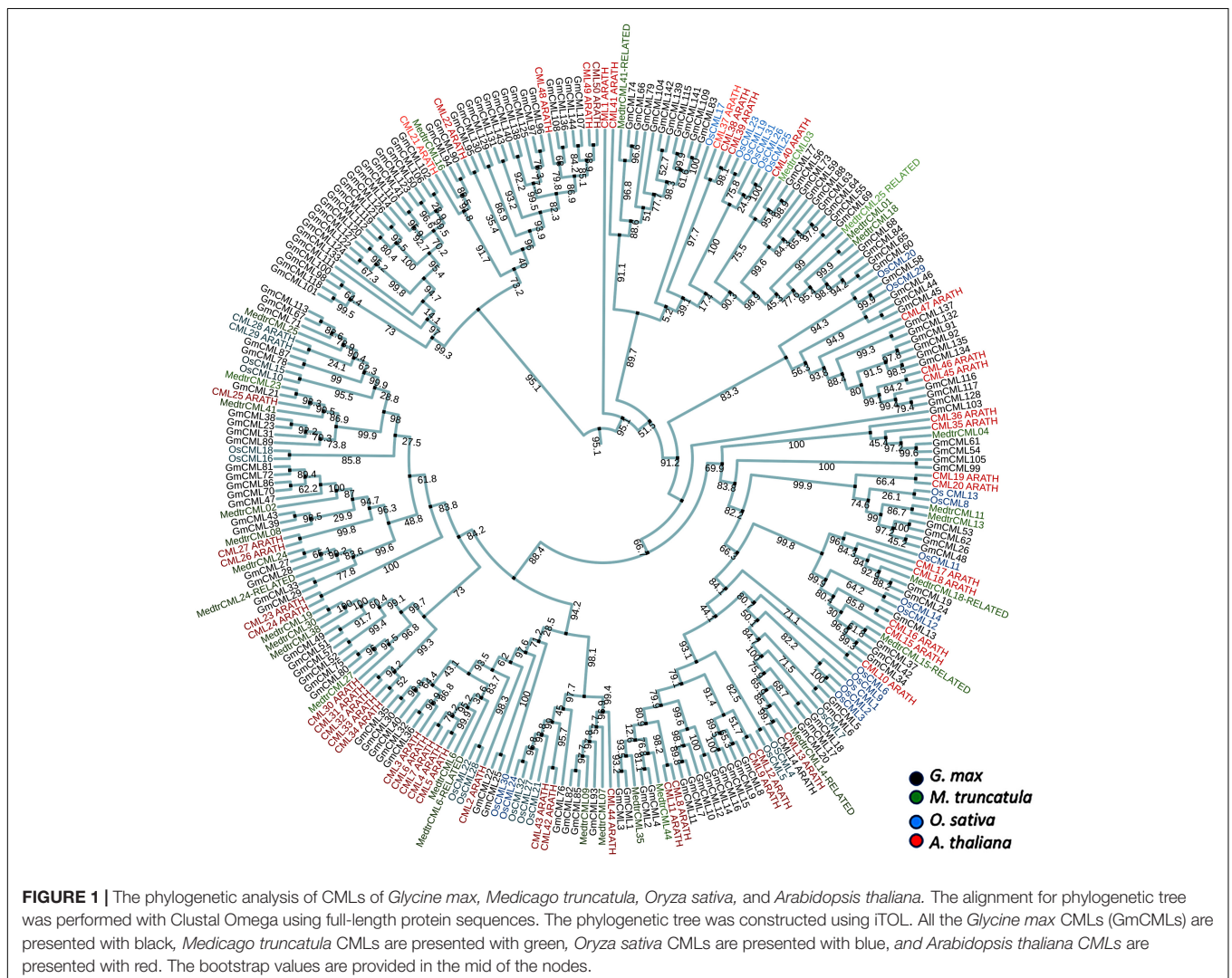
energy scores of docked molecules were compared with that of the Calmodulin 2 protein of *Arabidopsis thaliana* with Ca²⁺.

RESULTS

Genome-Wide Survey, Phylogenetic Study, and Similarity Search Identified True *Glycine max CMLs* in *Glycine max* Genome

A total of 144 CMLs were predicted in the soybean genome by Zeng et al. (2017) in a preliminary study. The protein sequences of these predicted CMLs were procured and named as GmCMLs. Phylogenetic analysis of GmCMLs with CMLs of other plants, including monocots and leguminous dicots was executed, which showed evolutionary relatedness among them. This analysis revealed that GmCML101, 118, 98, and 100 shared ancestry. GmCML113, 167, 71, and CML25 from *M. truncatula* and CML28 and CML29 from *A. thaliana* diverged recently

[21]https://pubchem.ncbi.nlm.nih.gov/

[22]http://www.cheminfo.org/Chemistry/Cheminformatics/FormatConverter/index.html

[23]https://bioinfo3d.cs.tau.ac.il/FireDock/

FIGURE 1 | The phylogenetic analysis of CMLs of *Glycine max, Medicago truncatula, Oryza sativa,* and *Arabidopsis thaliana*. The alignment for phylogenetic tree was performed with Clustal Omega using full-length protein sequences. The phylogenetic tree was constructed using iTOL. All the *Glycine max* CMLs (GmCMLs) are presented with black, *Medicago truncatula* CMLs are presented with green, *Oryza sativa* CMLs are presented with blue, *and Arabidopsis thaliana CMLs* are presented with red. The bootstrap values are provided in the mid of the nodes.

(**Figure 1**). Further, the GmCMLs were compared with CMLs of *A. thaliana*, revealing that 41 CMLs showed 50% or more sequence similarity with CMLs of *A. thaliana*. Sixteen of the GmCMLs showed higher sequence similarity (85–70%), whereas 25 GmCMLs showed more than 50% sequence similarity with CMLs of *A. thaliana*. The rest of the CMLs (103 GmCMLs) exhibited less than 50% homology, and 3 of them did not match with any of the CMLs of *A. thaliana*. With an identity of 50% as a threshold, 41 GmCMLs with 50% or more identities were labeled as true GmCMLs. A list of GmCMLs and their homologous CML of *A. thaliana* and their percentage similarity is presented in **Supplementary Table 1**.

Differential Expression of *Glycine max* *CMLs* During *Spodoptera litura-* Infestation and Wounding

To gain information on the involvement of *GmCMLs* in plant defense against *S. litura*-infestation, the expressions of true 41 *GmCMLs* were checked using qPCR with gene-specific primers (**Supplementary Table 2**), and the results indicated that 34
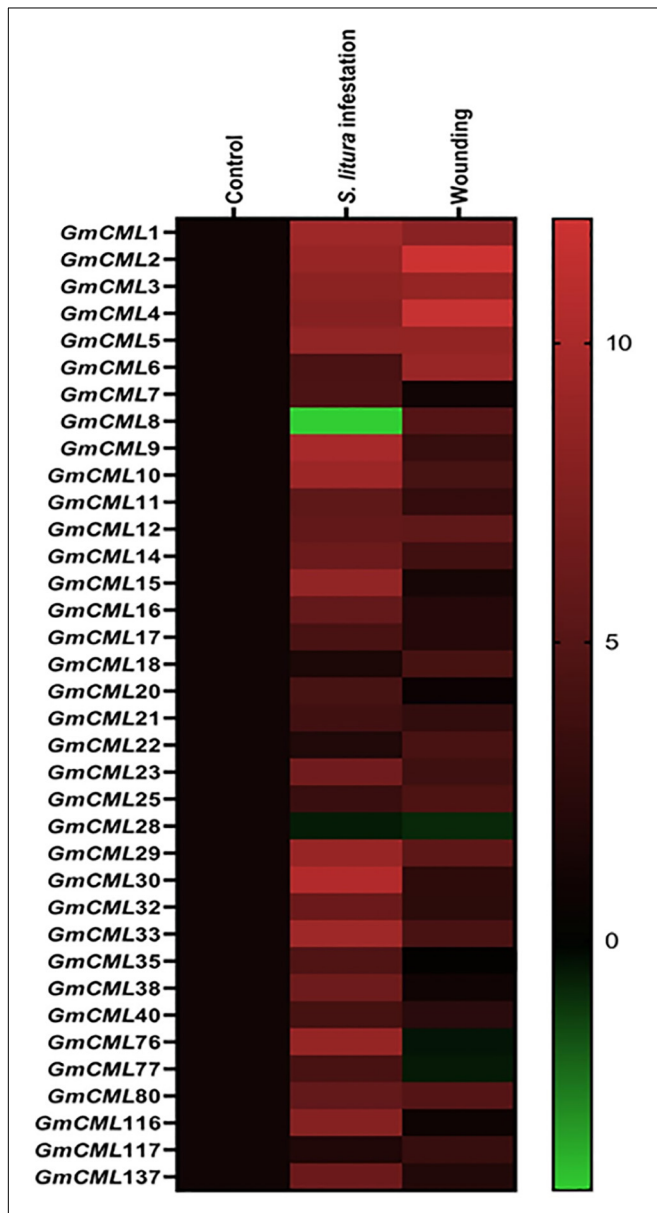

FIGURE 2 | Heat map showing expression profile of *GmCMLs* upon *S. litura*-infestation and wounding. The expression profile of *GmCMLs* upon *S. litura*-infestation and wounding is represented as heat map, generated with the help of Graphpad Prism software using $\log^2$ values of expression data obtained from qPCR. The qPCR data were analyzed utilizing $2^{-\Delta\Delta CT}$ method and statistical analysis was performed using ANOVA and Tukey's test ($P < 0.05$).
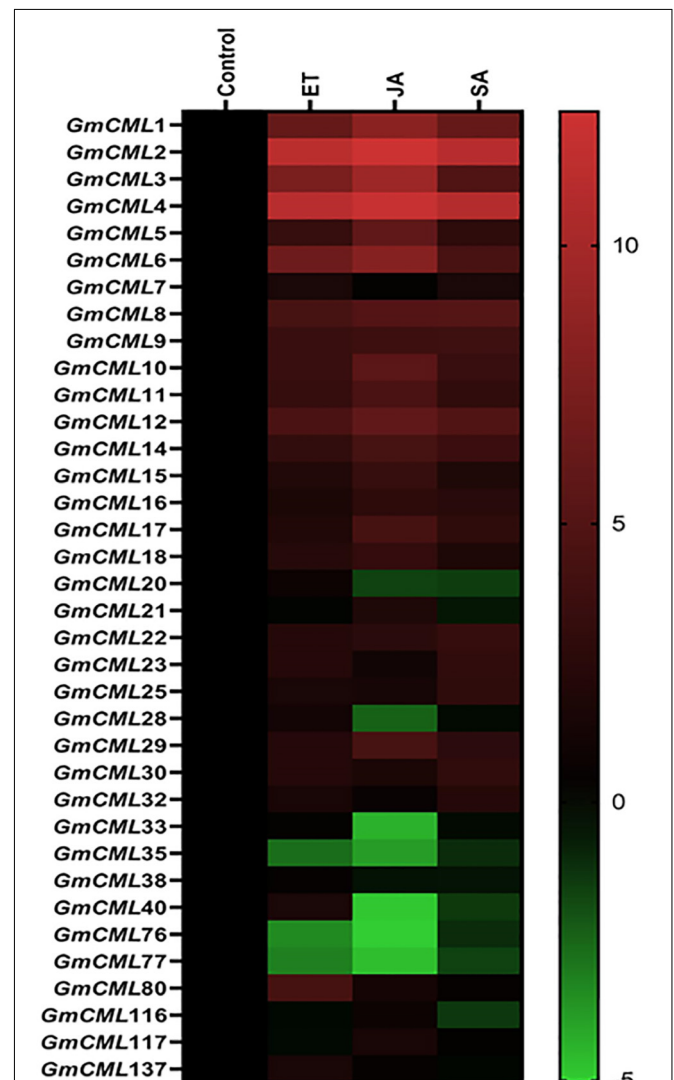

FIGURE 3 | Heat map showing expression profile of *GmCMLs* on application of JA, SA, and ET. The expression profile of *GmCMLs* on application of JA, SA, and ET is represented as a heat map, generated with the help of Graphpad Prism software using $\log^2$ values of expression data obtained from qPCR. The qPCR data were analyzed utilizing the $2^{-\Delta\Delta CT}$ method and statistical analysis was performed using ANOVA and Tukey's test ($P < 0.05$).

of the *GmCMLs* were upregulated while 2 of the *GmCMLs* were downregulated. *GmCML*30 showed a maximum change in expression while *GmCML*33, 9, 10, 1, and *GmCML*2, 76, 29, 3, 15, and 5 also showed a high level of expression (**Figure 2**; **Supplementary Figure 1**). The results demonstrated

that these *GmCMLs* were responsive to *S. litura*-infestation. However, a different transcript pattern was observed when plants were wounded/mechanically damaged; 24 of the *GmCMLs* were upregulated, and 5 *GmCMLs* were downregulated. Seven of the *GmCMLs* (*GmCMLs* 7, 20, 21, 35, 76, 77, and 116) exhibited

TABLE 1 | Gene locus, chromosome number, number of amino acids, number of EF-hands, molecular weight, pI, and subcellular localization of *S. litura*-inducible GmCMLs.

Gene name	Locus	Chromosome	Amino acid	Number of EF-hands	MW (kDa)	pI	PSORT subcellular localization
GmCML1	Glyma.10G178400	Chr10	150	4	17.01	4.04	chlo: 5, cyto: 5, extr: 2, cysk_nucl: 1.33333, cysk_plas: 1.33333
GmCML2	Glyma.02G002100	Chr02	150	4	17	3.94	cyto: 6.5, cyto_nucl: 4.5, chlo: 4, extr: 2, nucl: 1.5
GmCML3	Glyma.20G211700	Chr20	150	4	17	4.03	chlo: 5, cyto: 5, extr: 2, cysk_nucl: 1.33333, cysk_plas: 1.33333
GmCML4	Glyma.10G002200	Chr10	150	4	16.89	4.01	cyto: 7, cyto_nucl: 6.5, chlo: 4, extr: 1
GmCML5	Glyma.13G074800	Chr13	140	4	15.81	4.15	cyto: 6, nucl: 3, extr: 3, chlo: 1, golg: 1
GmCML6	Glyma.20G048900	Chr20	165	4	18.77	4.6	chlo: 12, mito: 2
GmCML7	Glyma.19G244300	Chr19	149	4	17.09	4.09	cyto: 10, nucl: 2, golg: 2
GmCML8	Glyma.19G160100	Chr19	148	4	16.7	4.17	cyto: 11, extr: 2, nucl: 1
GmCML9	Glyma.03G157800	Chr03	148	4	16.86	4.16	cyto: 14
GmCML10	Glyma.03G246800	Chr03	149	4	17.07	4.12	cyto: 10, nucl: 2, extr: 1, golg: 1
GmCML11	Glyma.19G098900	Chr19	114	2	13.25	4.15	cyto: 7, chlo: 3, nucl: 2, extr: 2
GmCML12	Glyma.10G161900	Chr10	149	4	17.21	4.32	cyto: 6, cyto_nucl: 5.83333, cyto_E.R.: 4, chlo: 3, nucl: 2.5, extr: 1, golg: 1
GmCML14	Glyma.20G224300	Chr20	149	4	17.2	4.35	cyto_nucl: 5.83333, cyto: 5, nucl: 3.5, cyto_E.R.: 3.5, chlo: 3, extr: 1, golg: 1
GmCML15	Glyma.02G143800	Chr02	149	4	17.19	4.28	cyto: 12, extr: 1, golg: 1
GmCML16	Glyma.10G030500	Chr10	149	4	17.12	4.25	cyto: 13, golg: 1
GmCML17	Glyma.11G030100	Chr11	147	4	16.55	4.73	cyto: 5, nucl_plas: 4, nucl: 3.5, plas: 3.5, chlo: 1, extr: 1
GmCML18	Glyma.01G211700	Chr01	147	4	16.54	4.75	cyto: 8, plas: 3, chlo: 1, nucl: 1, extr: 1
GmCML20	Glyma.17G134900	Chr17	147	4	16.55	5.66	cyto: 5, nucl: 2, plas: 2, extr: 2, pero: 1, cysk: 1, golg: 1
GmCML21	Glyma.05G238400	Chr05	188	4	20.43	4.28	nucl: 9, chlo: 3, mito: 1, extr: 1
GmCML22	Glyma.17G112000	Chr17	229	4	25.83	4.63	chlo: 9, cyto: 2, nucl: 1, extr: 1, E.R.: 1
GmCML23	Glyma.01G094000	Chr01	152	4	16.38	4.05	cyto: 6, chlo: 4, extr: 2, golg: 1, cysk_nucl: 1
GmCML25	Glyma.13G159600	Chr13	229	4	25.58	4.75	chlo: 7, extr: 3, vacu: 2, cyto: 1, cysk_nucl: 1
GmCML28	Glyma.11G217200	Chr11	137	4	15.49	5.07	mito: 8, nucl: 3, chlo: 1, cyto: 1, extr: 1
GmCML29	Glyma.14G215800	Chr14	141	4	15.89	4.62	cyto: 6, mito: 6, nucl: 1, extr: 1
GmCML30	Glyma.03G127000	Chr03	152	4	17	4.32	nucl: 4, chlo: 3, cyto: 3, mito: 2, extr: 2
GmCML32	Glyma.11G127500	Chr11	106	3	12.03	4.12	chlo: 3, extr: 3, cysk_nucl: 3, nucl: 2.5, cysk: 2.5, cyto: 2, golg: 1
GmCML33	Glyma.02G245700	Chr02	141	4	15.89	4.62	mito: 9, chlo: 3, nucl: 1, extr: 1
GmCML35	Glyma.19G129800	Chr19	152	4	16.95	4.25	nucl: 5, extr: 4, chlo: 2, ito: 2, cyto: 1
GmCML38	Glyma.09G067600	Chr09	102	2	11.26	4.08	cyto: 7, chlo: 3, extr: 2, nucl: 1, cysk: 1
GmCML40	Glyma.13G344200	Chr13	184	4	21.26	4.98	chlo: 9, mito: 3, cyto: 2
GmCML76	Glyma.09G270900	Chr09	183	3	20.52	4.48	mito: 8, chlo: 4, nucl: 2
GmCML77	Glyma.16G059300	Chr16	140	4	15.96	4.49	nucl: 6, chlo: 5, extr: 2,plas: 1
GmCML80	Glyma.07G212000	Chr07	185	4	21.2	6.84	chlo: 5, mito: 5, nucl: 4
GmCML116	Glyma.09G236800	Chr09	207	2	24.06	4.44	chlo: 5, nucl: 4, extr: 2, cyto: 1, cysk: 1, golg: 1
GmCML117	Glyma.18G260700	Chr18	207	3	24.08	4.51	chlo: 5, nucl: 5, extr: 2, cyto: 1, golg: 1
GmCML137	Glyma.04G078400	Chr04	178	4	19.71	4.61	nucl: 5, cyto: 5, chlo: 3, extr: 1

a significant change in mRNA expression specifically during *S. litura*-infestation.

Differential Expression of *Glycine max CMLs* Upon Foliar Application of Signaling Compounds

To recognize the role of defense-related signaling compounds in the regulation of *S. litura* inducible-*GmCMLs*, expression analysis was performed with plants treated with JA, SA, and ethylene (ET) by qPCR. *GmCMLs* 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, and 12 showed induced expression in all four treatments. The JA application also induced the expression of *GmCMLs* 14, 17, and 29. The ethylene upregulated the expression of 13 of the *GmCMLs,* and eight of them were commonly upregulated by JA and ethylene. In addition, eleven of the *GmCMLs* were upregulated by SA (**Figure 3**; **Supplementary Figure 2**).

Subcellular Localization, Molecular Weight, Isoelectric Point, and EF-Hand Analysis of *Glycine max* CMLs

The subcellular localization study using *in silico* tools predicted GmCMLs to be located mainly in the nucleus, cytosol, or chloroplast. Only a few of them were predicted to be located on the plasma membrane. The GmCMLs ranged from 229 (GmCML22) to 102 (GmCML38) amino acids, and their molecular weight ranged between 11.26 and 25.85 KDa. GmCML38, 32, and 11 were smallest by molecular weights (11.26 KDa), whereas GmCMLs 22 and 25 were of the largest molecular weight (25.85 KDa). The isoelectric point values of the GmCMLs ranged from pI 3.94 to 6.84, where GmCML2 had a minimum isoelectric point (pI 3.94) and GmCML80 had a maximum isoelectric point (pI 6.84) (**Table 1**).

GmCMLs Are Largely Intronless/Carry 1–3 Introns and Share EF-Hand Signature Sequence and Other Common Motifs

The intron and exon display analysis of *GmCMLs* revealed that a larger number of these CMLs were intronless. Among the studied 36 *GmCMLs*, 22 *GmCMLs* coding genes were intronless, whereas the rest of the *GmCMLs* had one to three introns. *GmCMLs* 1, 2, 3, 4, 7, 8, 9, 10, 14, 15, and 16 have four exons and three introns (**Figure 4**). Additionally, prediction of the conserved domain revealed that a couple of motifs were conserved among the members of GmCML family proteins. Many of the GmCMLs (30 GmCMLs) had four EF-hands but GmCML 32, 76, and 117 had three EF-hands and GmCML11, 38, and 137 had only two EF-hands (**Figure 5**). All GmCML proteins contained the EELKEAFKVF**DKDGBGYISASE**LRHVMRSLGEKLTDEEVEQ, EAFSLF**DKBGDGCITVEE**LATIJRSLGQNPTEEE, and MIKEA **DLDGDGQVBYEE**FVKM motif confirming the presence of EF-hand signature sequence.
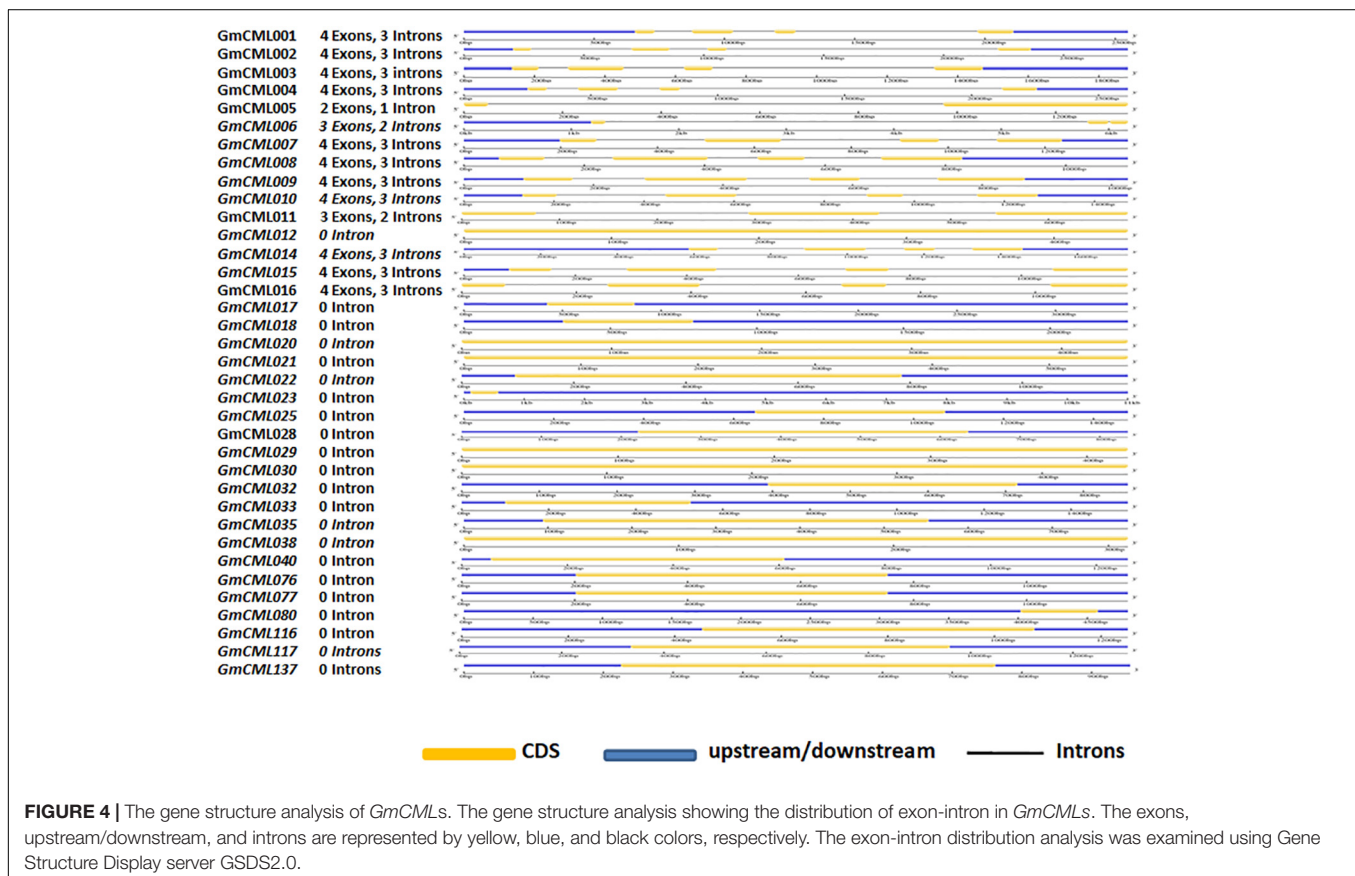

FIGURE 4 | The gene structure analysis of *GmCML*s. The gene structure analysis showing the distribution of exon-intron in *GmCML*s. The exons, upstream/downstream, and introns are represented by yellow, blue, and black colors, respectively. The exon-intron distribution analysis was examined using Gene Structure Display server GSDS2.0.
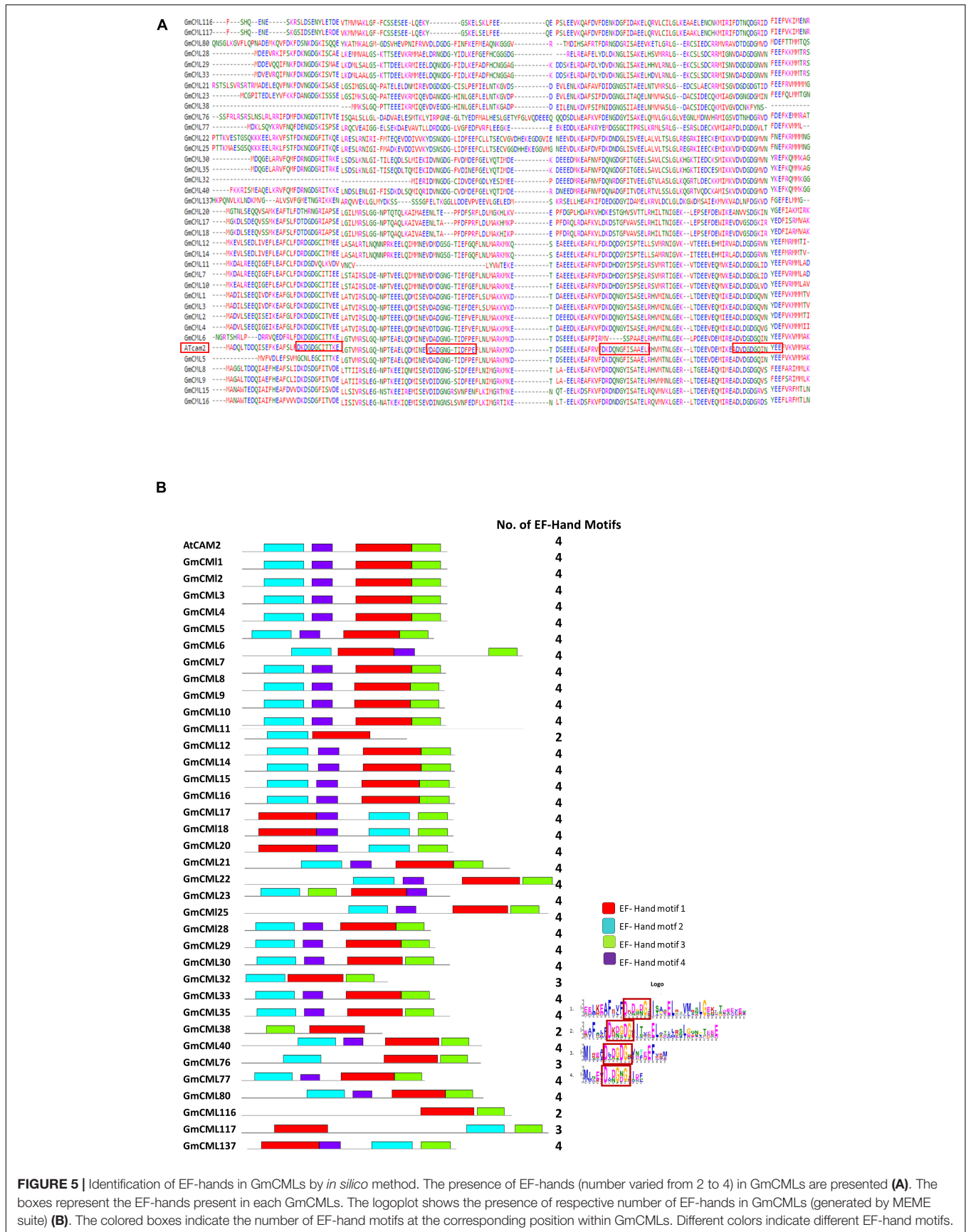
FIGURE 5 | Identification of EF-hands in GmCMLs by *in silico* method. The presence of EF-hands (number varied from 2 to 4) in GmCMLs are presented **(A)**. The boxes represent the EF-hands present in each GmCMLs. The logoplot shows the presence of respective number of EF-hands in GmCMLs (generated by MEME suite) **(B)**. The colored boxes indicate the number of EF-hand motifs at the corresponding position within GmCMLs. Different colors indicate different EF-hand motifs.

Members of *Glycine max CMLs* Family Differ in Their Tissue-Specific Expression Patterns

To identify the genes involved in distinct regulatory functions controlled by specific tissue, tissue-specific gene expression data were procured. The heat map was generated using RNA-seq data to examine the tissue-specific expression profile at the basal level (**Figure 6**; **Supplementary Table 3**). This analysis revealed that the relative expression level of *GmCMLs* 5, 6, and 40 was higher in all tissue types. *GmCMLs* 6, 5, 40, 28, 17, 80, 1, 4, and 29 showed their higher expression in stem. *GmCMLs* 5, 6, and 17 showed expression in seed and *GmCMLs* 6, 5, 40,

17, 80, and 137 expressed more in the nodule. *GmCML*17 had moderate expression in all the tissues, including flower, young leaf, green pods, root, nodule, stem, and seed. *GmCMLs* 6, 5, 40, 28, 17, 80, 1, 4, 29, 33, 76, and 32 exhibited maximum expression in root. *GmCMLs* 1, 6, 5, 40, 28, 17, and 80 exhibited higher expression in leaf. The mRNA of *GmCMLs* 6, 5, 40, 28, 17, and 80 were abundantly present in flower and pod. However, the rest of the genes showed minimal expression/no expression in different tissues.

Members of *Glycine max CML* Family Are Putative Targets of Known microRNAs

The miRNA target analysis (using psRNATarget) revealed 33 *GmCML* transcripts as putative targets of 125 miRNAs (**Table 2**).
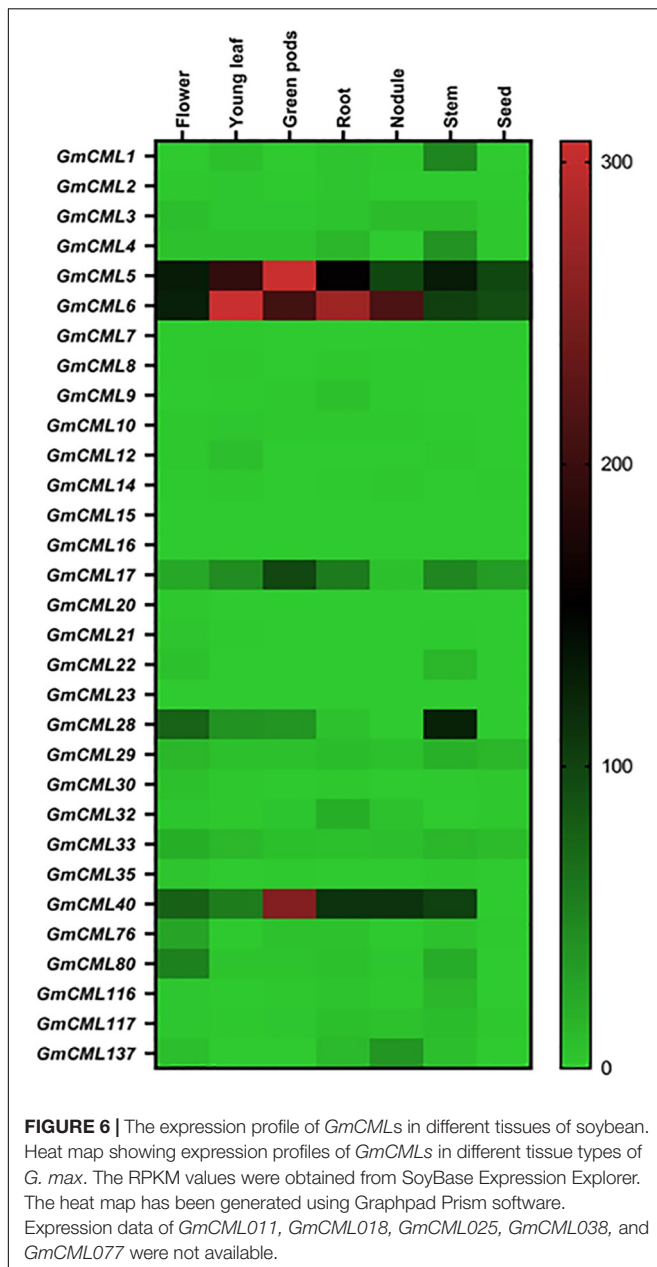

FIGURE 6 | The expression profile of *GmCMLs* in different tissues of soybean. Heat map showing expression profiles of *GmCMLs* in different tissue types of *G. max*. The RPKM values were obtained from SoyBase Expression Explorer. The heat map has been generated using Graphpad Prism software. Expression data of *GmCML011, GmCML018, GmCML025, GmCML038,* and *GmCML077* were not available.

TABLE 2 | Identification of miRNA target sites in *GmCML*-transcripts of *Glycine max.*

miRNA_Acc.	Target_Acc.	Number of miRNA targeting target gene
gma-miR171k-5p	Glyma.01G211700	3
gma-miR4352b	Glyma.02G002100	3
gma-miR1517	Glyma.02G143800	2
gma-miR159a-5p	Glyma.02G245700	3
gma-miR5376	Glyma.03G127000	1
gma-miR5667-5p	Glyma.03G157800	1
gma-miR167a	Glyma.03G246800	10
gma-miR172a	Glyma.04G078400	12
gma-miR159a-5p	Glyma.05G238400	2
gma-miR172a	Glyma.06G079900	14
gma-miR4370	Glyma.07G169100	5
gma-miR396a-5p	Glyma.07G212000	11
gma-miR4400	Glyma.09G236800	4
gma-miR4397-3p	Glyma.09G270900	3
gma-miR4352b	Glyma.10G002200	3
gma-miR5667-5p	Glyma.10G030500	2
gma-miR408d	Glyma.10G178400	2
gma-miR2119	Glyma.11G030100	4
gma-miR1514b-5p	Glyma.12G052100	6
gma-miR1515a	Glyma.13G074800	4
gma-miR4347	Glyma.13G159600	2
gma-miR4368a	Glyma.13G344200	1
gma-miR159a-5p	Glyma.14G215800	3
gma-miR4347	Glyma.17G112000	3
gma-miR171k-5p	Glyma.17G134900	2
gma-miR159a-5p	Glyma.18G039500	2
gma-miR396f	Glyma.18G260700	3
gma-miR5040	Glyma.19G098900	1
gma-miR4384	Glyma.19G129800	2
gma-miR1517	Glyma.19G160100	3
gma-miR5040	Glyma.19G244300	2
gma-miR1515a	Glyma.20G048900	4
gma-miR6299	Glyma.20G211700	2

Members of *Glycine max CMLs* Family Have an Overlapping Protein Network

To deduce the complex interplay of GmCMLs with other proteins, network and enrichment analysis of 36 selected GmCMLs was performed using the STRING database. The studies revealed that each GmCML corresponded to a particular string Id. Based on the STRING analysis, a protein protein interaction (PPI) network complex was constructed. The expected number of edges of the interaction network was 6, the average node degree of the network was 4, the average local clustering coefficient was 0.986, and the protein enrichment *p*-value was < 1.0e-16. The predicted results showed 40 nodes and 80 edges. Out of 36, 25 of the GmCMLs had shown their involvement as calcium-binding protein having EF-hand, where GmCML1, GmCML2, and GmCML15 had shown very high similarity with SCAM-4, SCAM-5, and Calmodulin, respectively, which were reported to be involved in plant-defense against pathogen attack (Park et al., 2004). Interestingly, the studies have shown the involvement of GmCML5 and GmCML6 in nodulation of soybean roots. Most of the GmCMLs showed to interact *via* GLYMA07G38930.2 (spindle-like microcephalyassociated protein isoform) and GLYMA17G01810.2 (spindle-like microcephalyassociated protein isoform). Overall, it was predicted that 24 GmCMLs have their role in plant-pathogen defense signaling and 5 of them showed their role in the phosphatidylinositol signaling and MAPK pathway (**Figure 7**).

In silico Promoter Analysis Identified Multiple Hormones and Stress-Responsive *Cis*-Elements in Promoter Sequences of *Glycine max CMLs*

A broad range analysis to decipher which *cis*-elements were present in the promoter regions of true *GmCMLs*, 1 kb 5′ UTR regions of each of the *GmCMLs* were subjected to Plant care analysis that predicted the presence of 16 different *cis*-regulatory elements involved in biotic or abiotic stresses. Most of the *GmCMLs* promoters were predicted to harbor more than one phytohormone-related *cis*-regulatory elements, such as ABA (ABRE), JA (G-box, Myc binding JA response element), SA (TCA-element), Ethylene (ERE), GA (P-Box and GARE-motif – gibberellin responsiveness), and biotic and abiotic stress-related *cis*-elements like ARE (Anaerobic induction), wounding, Myb-binding site involved in drought-inducibility (MBS), W-box-stress inducible and LTR (a low temperature)-responsive element, TC-rich repeats –defense and stress response (**Figure 8**).

Molecular Docking Indicates That Members of *Glycine max CML* Family Proteins Interact With Ca^{2+}

The study of the interaction of Ca^{2+} with GmCMLs could be crucial for understanding their mechanism of action. The tertiary structure of GmCMLs was modeled using Phyre^2, and the structures were validated by keeping a threshold value
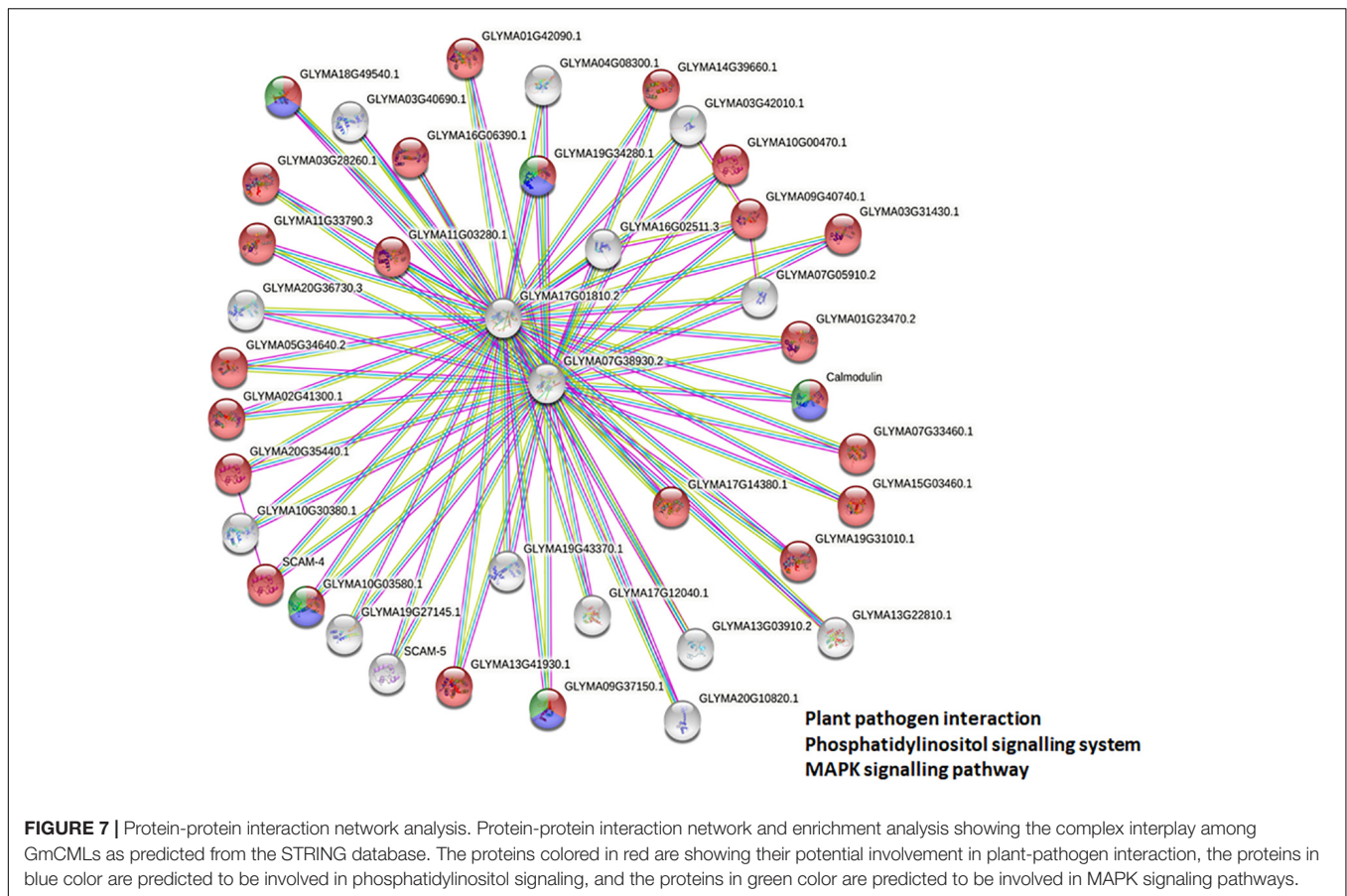
of >90% region of protein being present in the allowed region of the Ramachandran plot. The confidence level of modeled structures was more than 99%. The classic EF-hands consisting of a typical helix-loop-helix domain were detected and highlighted (**Figure 9**).

Previous studies had indicated that EF-hand motifs bind with Ca^{2+} for their activation. Thus, the molecular docking of 36 GmCMLs and Calmodulin 2 protein of *Arabidopsis thaliana* (AtCAM2) with Ca^{2+} were performed, and the interaction results of GmCMLs with Ca^{2+} were compared with the global energy score of AtCAM 2 with Ca^{2+}. The global energy score of AtCAM2 with Ca^{2+} was −2.94. The global energy score for 26 of the GmCMLs with Ca^{2+} ranged from −2.5 to −4.5. GmCMLs 17, 15, 29, 38, 116, 117, 33, 25, 30, 8, 35, 21, 76, 32, 77, 6, 22, 5, and 80 were showing the global energy scores from −4.41 to −3.05 whereas GmCMLs 20, 9, 12, 1, 10, 3, and 4 were showing global energy scores in between −2.8 and −2.5 (**Figure 9**).

DISCUSSION

Plants have to encounter numerous challenges ranging from biotic to abiotic factors during their lifespan. Therefore, they have evolved sophisticated defense mechanisms regulated by recognizing the specific type of stress and stimulation of certain early signaling events to defend themselves. One such crucial early event is the generation of Ca^{2+} ion pulses due to the elevation of influx and efflux, which acts as a cellular signal in response to external stimulus (Vadassery et al., 2012a,b). Calcium sensor proteins, such as CMLs, are essential components of Ca^{2+} signaling (DeFalco et al., 2010). We have only a little information about the involvement of CMLs in plant defense against herbivory in crop plants. Although fewer studies on the EF-hand possessing protein families, CaM, CDPK, CMLs, and CBLs (Weinl and Kudla, 2009; DeFalco et al., 2010; Poovaiah et al., 2013) have been taken into consideration in a couple of plants, such as *Arabidopsis thaliana, Oryza sativa,* Grapevine, *Nicotiana benthamiana,* and *Solanum lycopersicum* but significantly less information are available about CMLs in legumes (McCormack and Braam, 2003; Asano et al., 2005; Boonburapong and Buaboocha, 2007; Chen et al., 2013; Kong et al., 2013; Kleist et al., 2014).

Therefore, to characterize the CMLs of soybean and figure out their role in plant defense, the GmCMLs were subjected to phylogenetic analysis with CMLs of *Medicago truncatula, Arabidopsis thaliana,* and *Oryza sativa,* which revealed that many of the GmCMLs were grouped with CMLs of *Arabidopsis* and *Medicago*. They shared ancestry, which confirmed their evolutionary relationship. However, when GmCMLs were subjected to the BLAST search engine of TAIR against the known EF-hand protein of *A. thaliana*, only 41 GmCMLs had more than 50% sequence similarity. Expression analysis of true *GmCMLs* during *S. litura*-infestation was investigated, and the results demonstrated that many of these *GmCMLs* (36) were responsive to herbivory. Out of 41 genes, 34 of the *GmCMLs* were upregulated, while 2 of the *GmCMLs* were downregulated. Previous studies have reported that *CMLs* control signaling

FIGURE 7 | Protein-protein interaction network analysis. Protein-protein interaction network and enrichment analysis showing the complex interplay among GmCMLs as predicted from the STRING database. The proteins colored in red are showing their potential involvement in plant-pathogen interaction, the proteins in blue color are predicted to be involved in phosphatidylinositol signaling, and the proteins in green color are predicted to be involved in MAPK signaling pathways.

events upregulated during mechanical wounding and insect attack. It was observed that in *A. thaliana, CML9, 11, 12, 16, 17, 23,* and *42* genes were upregulated upon treatments with the oral secretion of larvae of the herbivorous insect *S. littoralis* (Vadassery et al., 2012a). In comparison to herbivory, 24 *GmCMLs* were upregulated 5 *GmCMLs* were downregulated by mechanical damage, and 8 of the *GmCMLs* showed induced expression exclusively during *S. litura*-infestation. These data suggested that wounding and herbivory had a different transcript pattern on account of the presence of insect-derived elicitors in the oral secretion during herbivory. The distinction in the transcript patterns during mechanical damage and insect attack has been detected in previous studies (Reymond et al., 2000, 2004; Singh et al., 2008). Wounding stimulated plant defense similar to herbivory; however, mechanical damage-induced responses were not the same as those activated by herbivory.

Signaling compounds, such as JA, ET, and SA, function as essential plant defense compounds against herbivory. Most of the herbivores trigger the jasmonate/ethylene pathway and the salicylate pathway (Hammerschmidt and Smith-Becker, 1999; Reymond et al., 2004). To recognize the role of defense regulators, change in the expression of *GmCMLs* was checked upon treatments with JA, SA, and ET. JA changed the mRNA levels of 14 of the *S. litura*-inducible *GmCMLs,* demonstrating the involvement of JA in the regulation of expression of *GmCMLs*.

JA-responsive *cis*-regulatory elements, TGACG and CGTCA, were observed in the promoter regions of *GmCMLs* 1, 3, 9, 14, 7 and 1, 3, 9, 14, 17, respectively, which further confirmed their regulation by JA. Ethylene upregulated the expression of 13 of the *GmCMLs,* and eight of them were commonly upregulated by JA. ET and JA were reported to induce defense genes synergistically in *Arabidopsis* (Kessler and Baldwin, 2002; Singh et al., 2008). In addition, eleven of the *GmCMLs* were upregulated by SA, indicating the role of SA in regulation of expression of *GmCMLs* either directly or through a crosstalk. The presence of SA-responsive *cis*-regulatory element in the promoter of *GmCML* 8 further confirmed their regulation by SA.

The exon-intron distribution study predicted the presence of zero to three introns in *GmCMLs*, indicating critical evolutionary changes in the *G. max* genome. There were only 11 *GmCMLs* that carried 1–3 introns, and 22 *GmCMLs* were intronless. The presence of few/no introns indicated their ability to get transcribed quickly to facilitate early defense response in the host plant during stress (Keshan et al., 2021).

The subcellular localization studies predicted the presence of CMLs in the nucleus, cytosol, and chloroplast mainly. Previous studies on CMLs had reported their localization in the nucleus and cytosol in *Arabidopsis* (Inzé et al., 2012; Vadassery et al., 2012a). The lower values of isoelectric points revealed that *GmCMLs* could be acidic, making them highly hydrophilic. The
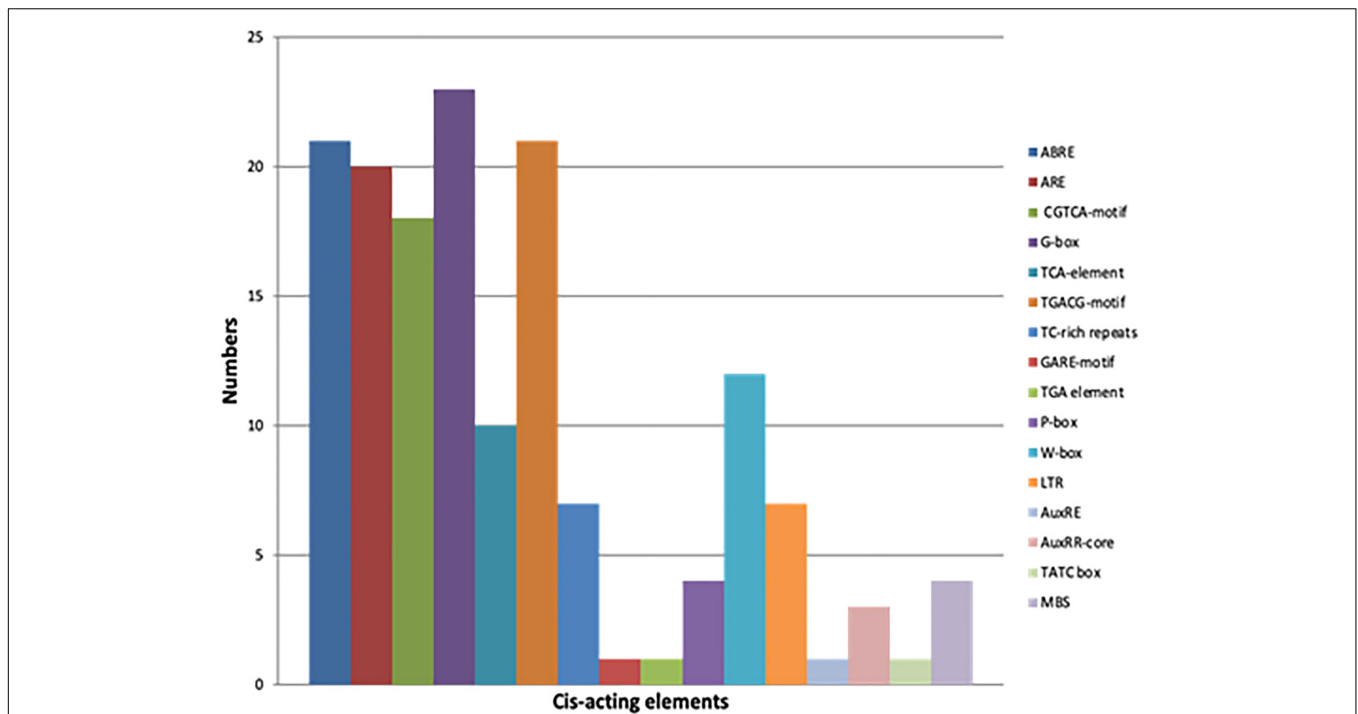
FIGURE 8 | The graphical representation of different *cis*-acting elements present in the promoter sequences of *GmCMLs*. Different colors depict different *cis*-acting elements [ABRE – Abscisic acid responsiveness, ARE –Anaerobic induction (wounding), CGTCA motif – Jasmonate responsiveness (MeJA), TCA-element – Salicylic acid responsiveness, TGACG-motif - Jasmonate responsiveness (MeJA), TC-rich repeats –defense and stress response, GARE-motif – Gibberellin responsiveness, TGA element – Auxin response, P Box - Gibberellin responsiveness, W-Box - Stress inducible, LTR – Low temperature response, AuxRE- Auxin responsive element, AuxRR-core - Auxin responsive element, TATC box - Gibberellin responsiveness and MBS – Drought inducible].

motif analysis exhibited 1 to 4 EF-hand potential motifs in most of the GmCMLs, indicating their potential identity as calcium-binding proteins (Boonburapong and Buaboocha, 2007; Kong et al., 2013; Kleist et al., 2014; Munir et al., 2016). Further, identifying potential miRNA target sites in *GmCMLs* divulged the dynamic roles of miRNA in post-transcriptional regulation of these genes in response to normal or stressed conditions (Bhatia et al., 2019).

The promoter of every gene possesses regulatory elements that act as binding sites for transcription factors. These promoter elements govern the differential expression of a gene in a tissue-specific manner at different growth stages and during environmental stresses. Several studies have divulged a direct connection between gene expression and promoter elements present in their 5′ upstream region (Lata et al., 2014). Therefore, *cis*-elements present in the promoter regions of all true *GmCMLs* were investigated using *in silico* tool, which predicted the presence of 16 different *cis*-regulatory elements involved in biotic or abiotic stresses, indicating their role in plant responses during stress and their involvement in signal transduction *via* ABA, SA, and JA.

The analysis of measuring mRNA abundance of genes belonging to a family in a tissue-specific manner will allow in identifying the genes involved in the development or regulatory pathways associated with a particular tissue type. The expression data of the *GmCML* gene family in different organs

exhibited divergence in their function. The transcript profiling of *GmCMLs* during several developmental stages in different organs revealed that most *GmCMLs* do not express vegetatively. They are inducible to exhibit a particular response, whereas some *GmCMLs*, such as *GmCML5*, *GmCML6*, and *GmCML40*, expressed highly in flower, root, pod, pod shell, young leaf, and seed, depicting their role in growth and development. Nine of the *GmCMLs* exhibited higher expression in the stem, illustrating their potential role in shoot development. Three of the *GmCMLs* showed their expression in seed, and six showed their expression in the nodule, indicating their prospective role in seed and nodule formation. Twelve *GmCMLs* showed higher expression in root, signifying their possible function in root development and plant-microbe interaction in soil. Seven of the *GmCMLs* exhibited higher expression in the leaf, specifying their role in leaf development and plant defense against pests and pathogens. Six of the *GmCMLs* were expressed abundantly in flower and pod, depicting their role in reproduction.

To deduce the complex interplay of GmCMLs with other proteins, network and enrichment analysis of *S. litura*-inducible 36 GmCMLs were performed with the STRING database. The PPI network complex indicated that out of 36 GmCMLs, 24 GmCMLs were predicted to play a role in plant-pathogen defense signaling; 5 of the GmCMLs (GmCML16, GmCML8, GmCML15, GmCML116, and GmCML117) could be involved in the phosphatidylinositol signaling system

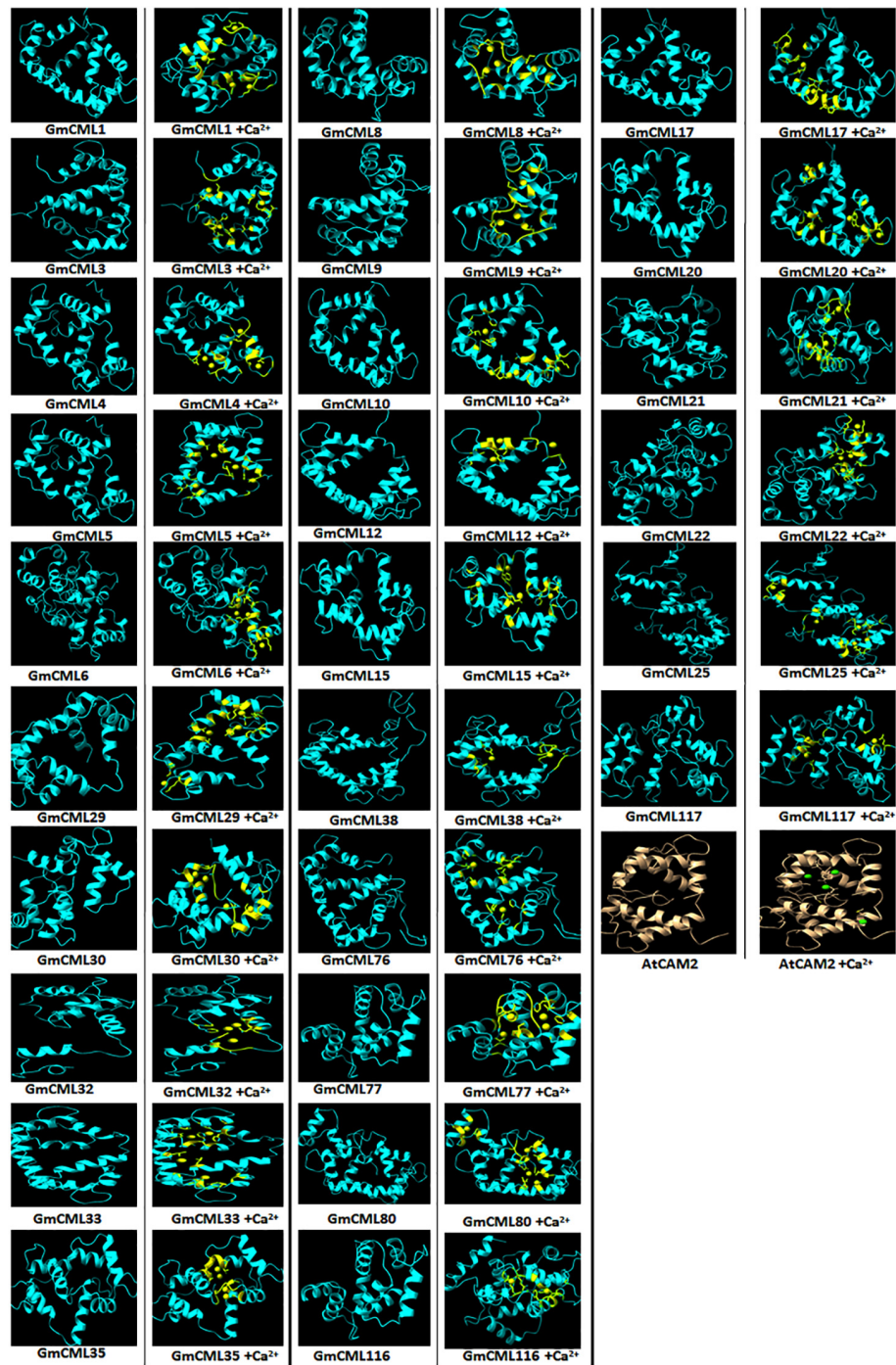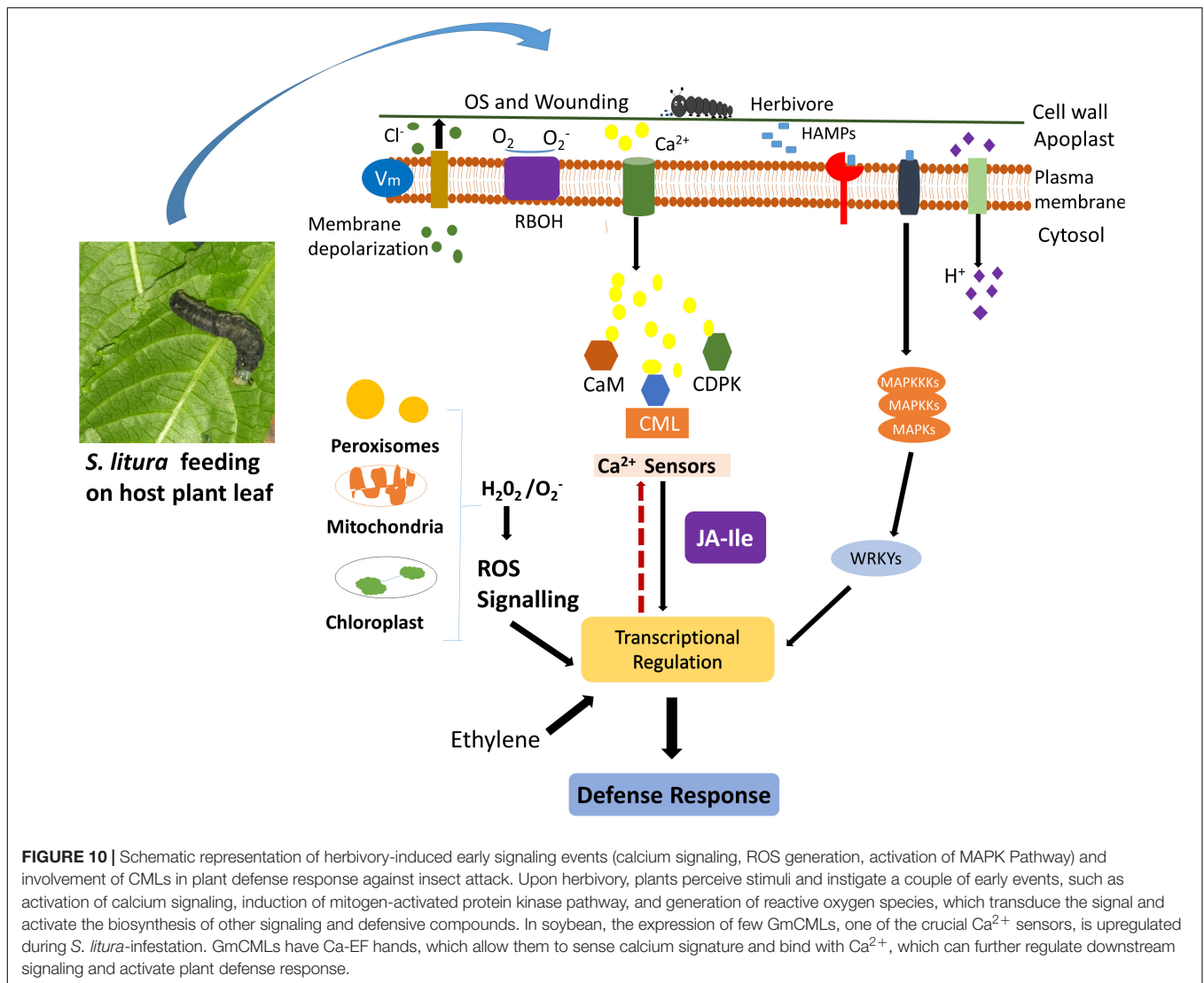**FIGURE 9 |** Interaction of GmCMLs with Ca²⁺ using molecular docking and their comparison with interaction of Calmodulin 2 (*A. thaliana*) with Ca²⁺. The figure shows the modeled tertiary structures of GmCMLs and their interaction with Ca²⁺, wherein the yellow balls indicate Ca²⁺ and the yellow-colored segments in the model represent EF-hands. The number of Ca²⁺ showing interaction with GmCMLs varied according to the number of EF-hands. The global docking energy score of interaction between 26 GmCMLs with Ca²⁺ is shown to have near to global docking energy score of interaction between Calmodulin 2 of *A. thaliana* with Ca²⁺. The tertiary structure of *A. thaliana* is shown in gray color and the green balls indicate Ca²⁺.

and MAPK signaling system. The plant defense signaling, phosphatidylinositol signaling system, MAPK pathway, and the components of these pathways, such as SCAM-4, SCAM-5, CaMEKK, CaZIK, CaRAF, PLA, PLC, and PLD had been reported to be involved in plant defense against biotic stresses (Park et al., 2004; Hung et al., 2014; Keshan et al., 2021). The above results are

FIGURE 10 | Schematic representation of herbivory-induced early signaling events (calcium signaling, ROS generation, activation of MAPK Pathway) and involvement of CMLs in plant defense response against insect attack. Upon herbivory, plants perceive stimuli and instigate a couple of early events, such as activation of calcium signaling, induction of mitogen-activated protein kinase pathway, and generation of reactive oxygen species, which transduce the signal and activate the biosynthesis of other signaling and defensive compounds. In soybean, the expression of few GmCMLs, one of the crucial Ca^{2+} sensors, is upregulated during *S. litura*-infestation. GmCMLs have Ca-EF hands, which allow them to sense calcium signature and bind with Ca^{2+}, which can further regulate downstream signaling and activate plant defense response.

prediction-based and should be perceived with caution. Further empirical analyses and experimentation would be required to assess whether these interactions are physiologically relevant.

The interaction of CMLs with Ca^{2+} is crucial for their biological activity. To examine the interaction between Ca^{2+} and GmCMLs, the tertiary structure of *S. litura*-inducible GmCMLs was predicted, which demonstrated a difference in the modeled structures of GmCMLs, mainly due to the change in their amino acid sequences. This diversity in their structures could allow them to perform a specific function. However, all the modeled structures of GmCMLs had shown the EF-hand helix-loop-helix motif as described in Calmodulin of *A. thaliana*, which could facilitate their interaction with Ca^{2+}. The molecular docking was performed to investigate the interaction of GmCMLs with Ca^{2+} and their global energy score of interaction was compared with the global energy score of interaction between AtCAM2 with Ca^{2+}. The results revealed that the global energy score for interaction between GmCMLs with Ca^{2+} ranged from -2.5 to -4.5 (for 26 of the GmCMLs), which is comparable to AtCAM2

with Ca^{2+} interaction. Thus, *in silico* interaction study predicted that calcium ions interacted with GmCMLs, particularly with EF-hand motifs, indicating that GmCMLs could be Ca^{2+}-binding proteins and could be involved in Ca^{2+} signaling. Further experimentation would be required to confirm whether CMLs are undergoing any conformational change to induce downstream signaling.

CONCLUSION

Plants encounter various environmental stresses, including insect attack during their life span. To defend themselves, plants have developed sophisticated signaling pathways that activate the biosynthesis of defensive proteins and metabolites. Among a couple of early events that instigate defense signaling, a transient change in cytosolic Ca^{2+} concentration is observed during herbivory. This transient change in Ca^{2+} concentration is identified by Ca^{2+} sensors, which transduce the information

to downstream target proteins leading to appropriate defense response. The role of Ca^{2+} sensors during herbivory has been studied in model plants, but not in leguminous plants, such as *G. max*. This study identified 41 true *GmCMLs*, an important Ca^{2+}-sensor, in soybean. The phylogenetic analysis revealed their similarity with already known *CMLs* from model plants. The presence of potential EF-hands further supported their identification as Ca^{2+}-sensors. The gene structure determination of *GmCMLs* indicated the presence of no/few introns revealing their capability to change their expression quickly during recognition of a stimulus. The differential expression of *GmCMLs* in various tissues signified their involvement in growth and development. Alteration in the transcript patterns of the novel *GmCMLs* during *S. litura*-infestation suggested their significant involvement in plant defense signaling during herbivory (**Figure 10**). Comparative expression profiling during mechanical damage showed that wounding-regulated responses are not the same as observed during herbivory. Signaling compounds, such as JA, ET, and SA, showed their involvement in regulating the expression of *S. litura*-inducible *GmCMLs*. Network and enrichment analysis of GmCMLs with other proteins indicated their role in plant defense, phosphatidylinositol signaling system, and MAPK signaling system. The study on the interaction between calcium and GmCMLs showed their identity as Ca^{2+}-binding proteins and their participation in signal transduction during Ca^{2+} signaling (**Figure 10**). The results of this study can be extended to confirm the role of calcium signaling in triggering the plant defense mechanism against *S. litura*-infestation. *S. litura*-inducible *GmCMLs* are prospective candidates for being critical components of calcium signaling during pest attacks. Thus, the structural and functional characterization of *S. litura*-inducible *GmCMLs* utilizing contemporary techniques like RNAi, overexpression, and/or gene editing is highly recommended to delineate the calcium signaling pathway and its role in plant defense during soybean-*S. litura* interaction.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

AS conceptualized and supervised the study. MY, JP, AS, IKS, AC, AR, MH, and JKK contributed to the investigation. MY, JP, and AS wrote – original draft preparation. AS, IKS, AC, AR, MH, and JKK contributed to writing, reviewing, editing, and visualization. AS, IKS, and MY contributed to formal analysis. AS, AC, and AR contributed to funding acquisition. IKS and AS contributed to resources. All authors have read and agreed to the published version of the manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.817950/full#supplementary-material

REFERENCES

Aldon, D., Mbengue, M., Mazars, C., and Galaud, J.-P. (2018). Calcium signalling in plant biotic interactions. *Int. J. Mol. Sci.* 19:665. doi: 10.3390/ijms19030665

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 Physiol.* 46, 356–366. doi: 10.1093/pcp/pci035

Badole, S. L., and Bodhankar, S. L. (2012). Glycine max (soybean) treatment for diabetes. *Bioact. Food Diet. Interv. Diabetes* 77, 77–82. doi: 10.1016/b978-0-12-397153-1.00008-1

Badole, S. L., and Mahamuni, S. P. (2013). "Soybean: Key Role in Skin Cancer," in *Bioactive Dietary Factors and Plant Extracts in Dermatology*, eds R. Watson and S. Zibadi (Totowa, N J: Humana Press), 315–320. doi: 10.1007/978-1-62703-167-7_28

Bender, K. W., and Snedden, W. A. (2013). Calmodulin-related proteins step out from the shadow of their namesake. *Plant Physiol.* 163, 486–495. doi: 10.1104/pp.113.221069

Bhatia, G., Sharma, S., Upadhyay, S. K., and Singh, K. (2019). Long non-coding RNAs coordinate developmental transitions and other key biological processes in grapevine. *Sci. Rep.* 9:3552. doi: 10.1038/s41598-019-38989-7

Boonburapong, B., and Buaboocha, T. (2007). Genome-wide identification and analyses of the rice calmodulin and related potential calcium sensor proteins. *BMC Plant Biol.* 7:4. doi: 10.1186/1471-2229-7-4

Chen, C.-Y., and Mao, Y.-B. (2020). Research advances in plant–insect molecular interaction. *Faculty Rev.* 9, F1000. doi: 10.12688/f1000research.21502.1

Chen, F., Fasoli, M., Tornielli, G. B., Dal Santo, S., Pezzotti, M., Zhang, L., et al. (2013). The evolutionary history and diverse physiological roles of the grapevine calcium-dependent protein kinase gene family. *PLoS One* 8:e80818. doi: 10.1371/journal.pone.0080818

Conrath, U., Pieterse, C. M. J., and Mauch-Mani, B. (2002). Priming in plant–pathogen interactions. *Trends Plant Sci.* 7, 210–216. doi: 10.1016/s1360-1385(02)02244-6

Cui, Z.-L., Gai, J.-Y., Ji, D.-F., and Ren, Z.-J. (1997). A study on leaf-feeding insect species on soybeans in Nanjing area. *Soybean Sci.* 16, 12–20.

DeFalco, T. A., Bender, K. W., and Snedden, W. A. (2010). Breaking the code: ca2+ sensors in plant signalling. *Biochem. J.* 425, 27–40. doi: 10.1042/BJ20091147

DeFalco, T. A., Moeder, W., and Yoshioka, K. (2016). Opening the gates: insights into cyclic nucleotide-gated channel-mediated signaling. *Trends Plant Sci.* 21, 903–906. doi: 10.1016/j.tplants.2016.08.011

DeFalco, T. A., and Zipfel, C. (2021). Molecular mechanisms of early plant pattern-triggered immune signaling. *Mol. Cell* 81, 3449–3467. doi: 10.1016/j.molcel.2021.07.029

Delk, N. A., Johnson, K. A., Chowdhury, N. I., and Braam, J. (2005). CML24, regulated in expression by diverse stimuli, encodes a potential Ca2+ sensor that functions in responses to abscisic acid, daylength, and ion stress. *Plant Physiol.* 139, 240–253. doi: 10.1104/pp.105.062612

Dobney, S., Chiasson, D., Lam, P., Smith, S. P., and Snedden, W. A. (2009). The calmodulin-related calcium sensor CML42 plays a role in trichome branching. *J. Biol. Chem.* 284, 31647–31657. doi: 10.1074/jbc.M109.056770

Dodd, A. N., Kudla, J., and Sanders, D. (2010). The language of calcium signaling. *Ann. Rev. Plant Biol.* 61, 593–620. doi: 10.1146/annurev-arplant-070109-104628

Fan, R., Wang, H., Wang, Y., and Yu, D. (2012). Proteomic analysis of soybean defense response induced by cotton worm (*Prodenia litura, fabricius*) feeding. *Proteome Sci.* 10, 16. doi: 10.1186/1477-5956-10-16

Gatehouse, J. A. (2002). Plant resistance towards insect herbivores: a dynamic interaction. *New Phytol.* 156, 145–169. doi: 10.1046/j.1469-8137.2002.00519.x

Gifford, J. L., Walsh, M. P., and Vogel, H. J. (2007). Structures and metal-ion-binding properties of the Ca2+-binding helix–loop–helix EF-hand motifs. *Biochem. J.* 405, 199–221. doi: 10.1042/BJ20070255

Gresshoff, P. M., and Ferguson, B. J. (2017). Molecular signals in nodulation control. *Int. J. Mol. Sci.* 18:125. doi: 10.3390/ijms18010125

Halitschke, R., and Baldwin, I. T. (2004). Jasmonates and related compounds in plant-insect interactions. *J. Plant Growth Regul.* 23, 238–245. doi: 10.1007/s00344-004-0037-z

Hammerschmidt, R., and Smith-Becker, J. A. (1999). "The role of salicylic acid in disease resistance," in *Mechanisms of Resistance to Plant Diseases*, eds A. Slusarenko, R. S. S. Fraser, and L. C. Van Loon (Kluwer: AcademicPublisher), 37–53.

Hancock, R. D., Hogenhout, S., and Foyer, C. H. (2015). Mechanisms of plant–insect interaction. *J. Exp. Bot.* 66, 421–424.

Heyer, M., Scholz, S. S., Reichelt, M., Kunert, G., Oelmüller, R., and Mithöfer, A. (2021). The Ca2+ sensor proteins CML37 and CML42 antagonistically regulate plant stress responses by altering phytohormone signals. *Plant Mol. Biol.* [Epub online ahead of print], doi: 10.1007/s11103-021-01184-2

Hu, B., Jin, J., Guo, A.-Y., Zhang, H., Luo, J., and Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31, 1296–1297. doi: 10.1093/bioinformatics/btu817

Hubbard, K., Hotta, C., Gardner, M., Braam, J., and Webb, A. (2008). The *Arabidopsis thaliana* calmodulin-like protein CML24 is a regulator of rhythmic Ca2+ signalling and flowering time. *Comp. Biochem. Physiol. Part A* 150:S153.

Hung, C.-Y., Aspesi, P. Jr., Hunter, M. R., Lomax, A. W., and Perera, I. Y. (2014). Phosphoinositide-signaling is one component of a robust plant defense response. *Front. Plant Sci.* 5:267. doi: 10.3389/fpls.2014.00267

Inzé, A., Vanderauwera, S., Hoeberichts, F. A., Vandorpe, M., Van Gaever, T. I. M., and Van Breusegem, F. (2012). A subcellular localization compendium of hydrogen peroxide-induced proteins. *Plant. Cell Environ.* 35, 308–320. doi: 10.1111/j.1365-3040.2011.02323.x

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. doi: 10.1038/nprot.2015.053

Keshan, R., Singh, I. K., and Singh, A. (2021). Genome wide investigation of MAPKKKs from *Cicer arietinum* and their involvement in plant defense against *Helicoverpa armigera*. *Physiol. Mol. Plant Pathol.* 115:101685. doi: 10.1016/j.pmpp.2021.101685

Kessler, A., and Baldwin, I. T. (2002). Plant responses to insect herbivory: the emerging molecular analysis. *Annu. Rev. Plant Biol.* 53, 299–328. doi: 10.1146/annurev.arplant.53.100301.135207

Kiep, V., Vadassery, J., Lattke, J., Maaß, J. P., Boland, W., Peiter, E., et al. (2015). Systemic cytosolic Ca2+ elevation is activated upon wounding and herbivory in Arabidopsis. *New Phytol.* 207, 996–1004. doi: 10.1111/nph.13493

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., et al. (2019). PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* 47, D1102–D1109. doi: 10.1093/nar/gky1033

Kleist, T. J., Spencley, A. L., and Luan, S. (2014). Comparative phylogenomics of the CBL-CIPK calcium-decoding network in the moss *Physcomitrella, Arabidopsis*, and other green lineages. *Front. Plant Sci.* 5:187. doi: 10.3389/fpls.2014.0018

Kliebenstein, D. J. (2004). Secondary metabolites and plant/environment interactions: a view through *Arabidopsis thaliana* tinged glasses. *Plant. Cell Environ.* 27, 675–684. doi: 10.1111/j.1365-3040.2004.01180.x

Kong, X., Lv, W., Jiang, S., Zhang, D., Cai, G., Pan, J., et al. (2013). Genome-wide identification and expression analysis of calcium-dependent protein kinase in maize. *BMC Genomics* 14:433. doi: 10.1186/1471-2164-14-433

Kumar, A., Panwar, R., Singh, A., and Singh, I. K. (2020). ""Role of Calcium Signalling During Plant–Herbivore Interaction", in *Plant Stress Biology*, eds B. Giri and M. P. Sharma (Singapore: Springer), 491–510. doi: 10.3390/cells10092219

Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8, 477–486. doi: 10.1007/BF00228148

Lata, C., Mishra, A. K., Muthamilarasan, M., Bonthala, V. S., Khan, Y., and Prasad, M. (2014). Genome-wide investigation and expression profiling of AP2/ERF transcription factor superfamily in foxtail millet (*Setaria italica* L.). *PLoS One* 9:e113092. doi: 10.1371/journal.pone.0113092

Leba, L., Cheval, C., Ortiz-Martín, I., Ranty, B., Beuzón, C. R., Galaud, J., et al. (2012). CML9, an Arabidopsis calmodulin-like protein, contributes to plant innate immunity through a flagellin-dependent signalling pathway. *Plant J.* 71, 976–989. doi: 10.1111/j.1365-313X.2012.05045.x

Lecourieux, D., Mazars, C., Pauly, N., Ranjeva, R., and Pugin, A. (2002). Analysis and effects of cytosolic free calcium increases in response to elicitors in *Nicotiana plumbaginifolia* cells. *Plant Cell* 14, 2627–2641. doi: 10.1105/tpc.005579

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van De Peer, Y., et al. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30, 325–327. doi: 10.1093/nar/30.1.325

Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301

Liao, J., Deng, J., Qin, Z., Tang, J., Shu, M., Ding, C., et al. (2017). Genome-wide identification and analyses of calmodulins and calmodulin-like proteins in *Lotus japonicas*. *Front. Plant Sci.* 8:482. doi: 10.3389/fpls.2017.00482

Lortzing, T., and Steppuhn, A. (2016). Jasmonate signalling in plants shapes plant–insect interaction ecology. *Curr. Opin. Insect Sci.* 14, 32–39. doi: 10.1016/j.cois.2016.01.002

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., et al. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 47, W636–W641. doi: 10.1093/nar/gkz268

Maffei, M., Bossi, S., Spiteller, D., Mithofer, A., and Boland, W. (2004).). Effects of feeding Spodoptera littoralis on lima bean leaves. I. Membrane potentials, intracellular calcium variations, oral secretions, and regurgitate components. *Plant Physiol.* 134, 1752–1762. doi: 10.1104/pp.103.034165

Magnan, F., Ranty, B., Charpenteau, M., Sotta, B., Galaud, J., and Aldon, D. (2008). Mutations in AtCML9, a calmodulin-like protein from *Arabidopsis thaliana*, alter plant responses to abiotic stress and abscisic acid. *Plant J.* 56, 575–589. doi: 10.1111/j.1365-313X.2008.03622.x

Mashiach, E., Schneidman-Duhovny, D., Andrusier, N., Nussinov, R., and Wolfson, H. J. (2008). FireDock: a web server for fast interaction refinement in molecular docking. *Nucleic Acids Res.* 36, W229–W232. doi: 10.1093/nar/gkn186

McCormack, E., and Braam, J. (2003). Calmodulins and related potential calcium sensors of *Arabidopsis*. *New Phytol.* 159, 585–598. doi: 10.1046/j.1469-8137.2003.00845.x

McCormack, E., Tsai, Y.-C., and Braam, J. (2005). Handling calcium signaling: *Arabidopsis* CaMs and CMLs. *Trends Plant Sci.* 10, 383–389. doi: 10.1016/j.tplants.2005.07.001

Meena, M. K., Prajapati, R., Krishna, D., Divakaran, K., Pandey, Y., Reichelt, M., et al. (2019). The Ca2+ channel CNGC19 regulates Arabidopsis defense against Spodoptera herbivory. *Plant Cell* 31, 1539–1562. doi: 10.1105/tpc.19.00057

Mithöfer, A., and Boland, W. (2008). Recognition of herbivory-associated molecular patterns. *Plant Physiol.* 146, 825–831. doi: 10.1104/pp.107.113118

Munir, S., Khan, M. R. G., Song, J., Munir, S., Zhang, Y., Ye, Z., et al. (2016). Genome-wide identification, characterization and expression analysis of calmodulin-like (CML) proteins in tomato (*Solanum lycopersicum*). *Plant Physiol. Biochem.* 102, 167–179. doi: 10.1016/j.plaphy.2016.02.020

Naresh, R. K., Dhaliwal, S. S., Chaudhary, M., Chandra, M. S., and Kumar, A. (2019). *Legumes: An Option to enhance Productivity and Soil Health Sustainability*. Singapore: Springer, 93.

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011). Open Babel: an open chemical toolbox. *J. Cheminform.* 3:33. doi: 10.1186/1758-2946-3-33

Park, H. C., Kim, M. L., Kang, Y. H., Jeon, J. M., Yoo, J. H., Kim, M. C., et al. (2004). Pathogen-and NaCl-induced expression of the SCaM-4 promoter is mediated in part by a GT-1 box that interacts with a GT-1-like transcription factor. *Plant Physiol.* 135, 2150–2161. doi: 10.1104/pp.104.041442

Park, H. C., Park, C. Y., Koo, S. C., Cheong, M. S., Kim, K. E., Kim, M. C., et al. (2010). AtCML8, a calmodulin-like protein, differentially activating CaM-dependent enzymes in *Arabidopsis thaliana*. *Plant Cell Rep.* 29, 1297–1304. doi: 10.1007/s00299-010-0916-7

Perochon, A., Aldon, D., Galaud, J.-P., and Ranty, B. (2011). Calmodulin and calmodulin-like proteins in plant calcium signaling. *Biochimie* 93, 2048–2053. doi: 10.1016/j.biochi.2011.07.012

Poovaiah, B. W., Du, L., Wang, H., and Yang, T. (2013). Recent advances in calcium/calmodulin-mediated signaling with an emphasis on plant-microbe interactions. *Plant Physiol.* 163, 531–542. doi: 10.1104/pp.113.220780

Ranty, B., Aldon, D., Cotelle, V., Galaud, J.-P., Thuleau, P., and Mazars, C. (2016). Calcium sensors as key hubs in plant responses to biotic and abiotic stresses. *Front. Plant Sci.* 7:327. doi: 10.3389/fpls.2016.00327

Reddy, A. S. N., Ben-Hur, A., and Day, I. S. (2011). Experimental and computational approaches for the study of calmodulin interactions. *Phytochemistry* 72, 1007–1019. doi: 10.1016/j.phytochem.2010.12.022

Reymond, P., Bodenhausen, N., Van Poecke, R. M. P., Krishnamurthy, V., Dicke, M., and Farmer, E. E. (2004). A conserved transcript pattern in response to a specialist and a generalist herbivore. *Plant Cell* 16, 3132–3147. doi: 10.1105/tpc.104.026120

Reymond, P., Weber, H., Damond, M., and Farmer, E. E. (2000). Differential gene expression in response to mechanical wounding and insect feeding in *Arabidopsis*. *Plant Cell* 12, 707–719. doi: 10.1105/tpc.12.5.707

Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H. J. (2005). PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* 33, W363–W367. doi: 10.1093/nar/gki481

Schuler, M. A. (2011). P450s in plant–insect interactions. *Biochim. Biophys. Acta* 1814, 36–45. doi: 10.1016/j.bbapap.2010.09.012

Singh, A., Kumar, A., Hartley, S., and Singh, I. K. (2020a). Silicon: its ameliorative effect on plant defense against herbivory. *J. Exp. Bot.* 71, 6730–6743. doi: 10.1093/jxb/eraa300

Singh, S., Singh, A., Kumar, S., Mittal, P., and Singh, I. K. (2020b). Protease inhibitors: recent advancement in its usage as a potential biocontrol agent for insect pest management. *Insect Sci.* 27, 186–201. doi: 10.1111/1744-7917.12641

Singh, S., Tyagi, C., Rather, I. A., Sabir, J. S. M., Hassan, M., Singh, A., et al. (2020c). Molecular modeling of chemosensory protein 3 from *Spodoptera litura* and its binding property with plant defensive metabolites. *Int. J. Mol. Sci.* 21:4073. doi: 10.3390/ijms21114073

Singh, A., Panwar, R., Mittal, P., Hassan, M. I., and Singh, I. K. (2021a). Plant cytochrome P450s: role in stress tolerance and potential applications for human welfare. *Int. J. Biol. Macromol.* 184, 874–886. doi: 10.1016/j.ijbiomac.2021.06.125

Singh, A., Singh, S., Singh, R., Kumar, S., Singh, S. K., and Singh, I. K. (2021b). Dynamics of *Zea mays* transcriptome in response to a polyphagous herbivore, Spodoptera litura. *Funct. Integr. Genom.* 21, 571–592. doi: 10.1007/s10142-021-00796-7

Singh, S., Singh, A., and Singh, I. K. (2021c). "Transcriptomics Studies Revealing Enigma of Insect-Plant Interaction," in *Plant-Pest Interactions: From Molecular Mechanisms to Chemical Ecology*, eds I. K. Singh and A. Singh (Singapore: Springer), 31–55. doi: 10.1007/978-981-15-2467-7_2

Singh, A., Singh, I. K., and Verma, P. K. (2008). Differential transcript accumulation in Cicer arietinum L. in response to a chewing insect *Helicoverpa armigera* and defence regulators correlate with reduced insect performance. *J. Exp. Bot.* 59, 2379–2392. doi: 10.1093/jxb/ern111

Singh, A., Singh, S., and Singh, I. K. (2016). Recent insights into the molecular mechanism of jasmonate signaling during insect-plant interaction. *Australas. Plant Pathol.* 45, 123–133. doi: 10.1007/s13313-015-0392-1

Singh, A., Tyagi, C., Nath, O., and Singh, I. K. (2018). Helicoverpa-inducible Thioredoxin h from *Cicer arietium*: structural modeling and potential targets. *Int. J. Biol. Macromol.* 109, 231–243. doi: 10.1016/j.ijbiomac.2017.12.079

Singh, I. K., and Singh, A. (2021). *Plant-Pest Interactions: From Molecular Mechanisms to Chemical Ecology*. New York NY: Springer.

Soyastats (2010). *A Reference Guide to Important Soybean Facts and Figure*. American Soybean Association.

Stotz, H. U., Pittendrigh, B. R., Kroymann, J., Weniger, K., Fritsche, J., Bauke, A., et al. (2000). Induced plant defense responses against chewing insects. Ethylene signaling reduces resistance of *Arabidopsis* against Egyptian cotton worm but not diamondback moth. *Plant Physiol.* 124, 1007–1018. doi: 10.1104/pp.124.3.1007

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49, D605–D612.

Thor, K., and Peiter, E. (2014). Cytosolic calcium signals elicited by the pathogen-associated molecular pattern flg22 in stomatal guard cells are of an oscillatory nature. *New Phytol.* 204, 873–881. doi: 10.1111/nph.13064

Vadassery, J., Reichelt, M., Hause, B., Gershenzon, J., Boland, W., and Mithöfer, A. (2012a). CML42-mediated calcium signaling coordinates responses to *Spodoptera* herbivory and abiotic stresses in *Arabidopsis*. *Plant Physiol.* 159, 1159–1175. doi: 10.1104/pp.112.198150

Vadassery, J., Scholz, S. S., and Mithöfer, A. (2012b). Multiple calmodulin-like proteins in *Arabidopsis* are induced by insect-derived (*Spodoptera littoralis*) oral secretion. *Plant Signal. Behav.* 7, 1277–1280. doi: 10.4161/psb.21664

Vandelle, E., Vannozzi, A., Wong, D., Danzi, D., Digby, A.-M., Dal Santo, S., et al. (2018). Identification, characterization, and expression analysis of calmodulin and calmodulin-like genes in grapevine (*Vitis vinifera*) reveal likely roles in stress responses. *Plant Physiol. Biochem.* 129, 221–237. doi: 10.1016/j.plaphy.2018.06.003

Vanderbeld, B., and Snedden, W. A. (2007). Developmental and stimulus-induced expression patterns of *Arabidopsis* calmodulin-like genes CML37 CML38 and CML39. *Plant Mol. Biol.* 64, 683–697. doi: 10.1007/s11103-007-9189-0

Vasudev, A., and Sohal, S. (2016). Partially purified Glycine max proteinase inhibitors: potential bioactive compounds against tobacco cutworm, *Spodoptera litura* (Fabricius, 1775)(Lepidoptera: Noctuidae). *Turkish J. Zool.* 40, 379–387. doi: 10.3906/zoo-1508-20

Vincent, T. R., Avramova, M., Canham, J., Higgins, P., Bilkey, N., Mugford, S. T., et al. (2017). Interplay of plasma membrane and vacuolar ion channels, together with BAK1, elicits rapid cytosolic calcium elevations in Arabidopsis during aphid feeding. *Plant Cell* 29, 1460–1479. doi: 10.1105/tpc.17.00136

Waese, J., Fan, J., Pasha, A., Yu, H., Fucile, G., Shi, R., et al. (2017). ePlant: visualizing and exploring multiple levels of data for hypothesis generation in plant biology. *Plant Cell* 29, 1806–1821. doi: 10.1105/tpc.17.00073

Weinl, S., and Kudla, J. (2009). The CBL–CIPK Ca2+-decoding signaling network: function and perspectives. *New Phytol.* 184, 517–528. doi: 10.1111/j.1469-8137.2009.02938.x

Zeng, H., Zhang, Y., Zhang, X., Pi, E., and Zhu, Y. (2017). Analysis of EF-hand proteins in soybean genome suggests their potential roles in environmental and nutritional stress signaling. *Front. Plant Sci.* 8:877. doi: 10.3389/fpls.2017.00877

Zhang, B. Q., and Yang, X. B. (2000). Pathogenicity of *Pythium* populations from corn-soybean rotation fields. *Plant Dis.* 84, 94–99. doi: 10.1094/PDIS.2000.84.1.94

Identification of Putative Elicitors From Plant Root Exudates Responsible for PsoR Activation in Plant-Beneficial *Pseudomonas* spp. by Docking and Molecular Dynamics Simulation Approaches to Decipher Plant–Microbe Interaction

Diksha Sati[1,2], Tushar Joshi[3], Satish Chandra Pandey[1], Veni Pande[1,3], Shalini Mathpal[3], Subhash Chandra[4]* and Mukesh Samant[1]*

[1]*Cell and Molecular Biology Laboratory, Department of Zoology (DST-FIST Sponsored), Soban Singh Jeena University Campus, Almora, India,* [2]*Department of Zoology, Kumaun University, Nainital, India,* [3]*Department of Biotechnology, Kumaun University, Sir J C Bose Technical Campus, Bhimtal, India,* [4]*Computational Biology & Biotechnology Laboratory, Department of Botany, Soban Singh Jeena University Campus, Almora, India*

Plants and rhizobacteria are coexisting since the beginning, but the exact mechanism of communication between them remains enigmatic. The PsoR protein of plant-beneficial *Pseudomonas* spp., a group of root-associated bacteria, is known to produce a range of antifungal and insecticidal secondary metabolites like 2,4-diacetyl phloroglucinol (DAPG), pyrrolnitrin, and chitinase making them great biocontrol agents and thus helping in plant growth promotion. To better understand the inter-kingdom signaling between plants and plant growth-promoting rhizobacteria (PGPR), the interaction of PsoR with various root exudates was investigated computationally. For this, we first modeled the PsoR protein and confirmed it using the Ramachandran plot. A total of 59 different low molecular weight phytochemicals, secreted as root exudates by plants, were identified by extensive text mining. They were virtually screened with the PsoR protein by molecular docking. Based on the lowest binding energy, ranging from -7.1 to $-6.3\,\text{kcal}\,\text{mol}^{-1}$, the top five exudates were chosen. To analyze the stability of the docked protein–ligand complex, a molecular dynamics (MD) simulation of 100 nanoseconds was done. Two root exudates, saponarin and 2-benzoxazolinone (BOA), showed suitable binding with PsoR by forming hydrogen, hydrophobic, and Van der Waals interactions. To confirm the MD simulation results, RMSF, RG, SASA, and interaction energy were calculated. This computational study first time reports that saponarin and 2-BOA, predominantly present in the root exudates of barley and wheat, respectively, demonstrate effective binding with the modeled PsoR protein and are likely of showing cross-kingdom interactions.

Keywords: inter-kingdom signaling, saponarin, 2-benzoxazolinone, molecular docking, molecular dynamic simulation

INTRODUCTION

Plants and rhizobacteria have been living nearby and co-evolving for a considerably long period, during which time both have been subjected to the signaling molecules formed and released by the other (Lyu et al., 2021). Rhizobacteria employ various direct and indirect mechanisms of growth promotion in their host plants (Papenfort and Bassler, 2016). Plants, in return, exude some active metabolites from their roots, of diverse chemical nature, and help in shaping microbial communities in the rhizosphere (Papenfort and Bassler, 2016; Sasse et al., 2018). Plant–microbe interaction not only maintains plant productivity but also ensures global food security by supporting a healthy ecosystem (Sati et al., 2022). Traditionally, LuxR protein has been known as an important regulator of quorum sensing (QS) signaling, by detecting the N-acyl homoserine lactones (AHLs) molecules in a cell density-dependent manner and regulating the target gene (Papenfort and Bassler, 2016). LuxR solos are present in various bacteria and play diverse roles like plant growth promotion, nodule formation, locomotion, extra-chromosomal DNA transfer, pathogenesis, and regulation of QS (Subramoni and Venturi, 2009). Moreover, LuxR solo of plant-associated bacteria (both plant-friendly and plant-pathogenic ones) differs in only one or two critical amino acid residues in the auto-inducer domain, which possibly could allow them to interact with plant-derived signal molecules instead of AHLs, hence facilitating inter-kingdom signaling (Subramoni et al., 2011; González et al., 2013). Various LuxR solo proteins have been reported, viz. XccR in Xanthomonas campestris, OryR in Xanthomonas oryzae, NesR in Sinorhizobium meliloti, and PsoR in plant-beneficial Pseudomonas spp. The PsoR protein of Pseudomonas fluorescens species complex is concerned with the regulation of biocontrol traits. The psoR gene codes for a 252-residue long protein containing an N terminal inducer-binding domain (PF03472) and a C terminal DNA-binding domain (PF00196), a characteristic of LuxR proteins involved in QS. PsoR directs transcriptional modulation of various antimicrobial-related genes in plant-beneficial Pseudomonas spp. in reaction to plant compounds (Patel et al., 2013). Pyrrolnitrin biosynthetic genes, DAPG genes, and genes involving synthesis of chitin-binding protein and chitinase were upregulated by PsoR (Haas and Keel, 2003). The interconnection between the PsoR protein of plant-beneficial Pseudomonas spp. and root exudates of rice and wheat plants has been previously reported (González et al., 2013). Further, the psoR gene in P. fluorescens CHA0 protects the wheat plant against Pythium ultimum infection because, when the psoR gene was knocked out from P. fluorescens, it was unable to protect the plant root, and a 30% decrease in fresh weight of root was seen in comparison with control. When this experiment was conducted taking cucumber as the host plant, though P. fluorescens CHA0 imparted protection from the pathogen, there was no considerable difference between the wild-type strain and the psoR mutant (Subramoni et al., 2011). These findings suggest that the psoR gene is likely specific for a molecule present in Poaceae plants. Tracing the given

lines of evidence, we designed an in silico study to determine the potential plant root exudates interacting with PsoR and further manifesting its role in plant growth promotion by biocontrol. By cutting time and expenditure and hastening studies on desired interactions with low molecular weight molecules, computational approaches have been highly useful tools in finding the function of any protein. Molecular docking has shown to be one of the most useful in silico strategies for identifying new bioactive compounds from large chemical libraries. During docking, the primary binding mode for each small molecule is analyzed inside a target binding pocket, and a docking score is assigned to them based on their binding propensity. It can also be used to assist lead refinement by showing how a certain hit can be altered to optimize protein–ligand interaction. In comparison to in vitro techniques, molecular docking is both cost-friendly and time-saving. However, many estimations are generated by docking procedures and majority of them absent receptor flexibility. As a result, the dependability of the resulting protein–ligand complexes is limited. Consequently, molecular docking must be paired with other in silico methods to obtain more accurate results. Combining docking with MD simulations is a prevalent approach to strengthen results. MD simulations work by improving the topologies of the already docked receptor–ligand complexes, providing more accurate interaction energies, and revealing the ligand-binding mechanism. Molecular modeling, docking, and dynamics simulation studies have been immensely helpful in the past for elucidating the structure and function of many plant proteins. For example, computational analysis of α-expansin protein from mountain papaya fruit was made possible using a comparative modeling strategy and its optimization by MD simulations (Gaete-Eastman et al., 2015). Similarly, structural insights of rice protein urease were provided by using homology modeling and MD simulation approach (Kumar et al., 2018). However, computational studies made on plant–microbe interaction are quite limited. Earlier, the ligand-binding sites of QS LuxRs were mapped using structure-based sequence alignment and structural superimposition (Covaceuszach et al., 2013). Bioinformatics and molecular docking-based approaches have previously been successfully applied for providing structural knowledge of regulatory domains of many Lux R family proteins, including PsoR (Karplus and McCammon, 2002; Trott and Olson, 2010; Laskowski et al., 2018). However, more focused docking and simulation-based studies were required to figure out the exudate compounds eliciting bacterial protein PsoR. Hence, in our study, a total of 59 diverse plant root exudate compounds were taken and those showing strong interaction with PsoR were selected by the computational screening method (**Figure 1**). Using an in silico study, we found two putative root exudates, viz. saponarin and 2-benzoxazolinone (BOA), positively interact with PsoR of P. fluorescens. Saponarin is the major flavonoid found in barley, while BOA is the major secondary metabolite of wheat, maize, and rye. We also explored PsoR catalytic sites and using molecular dynamics (MD) simulation authenticated the reliability of PsoR-exudate complexes.
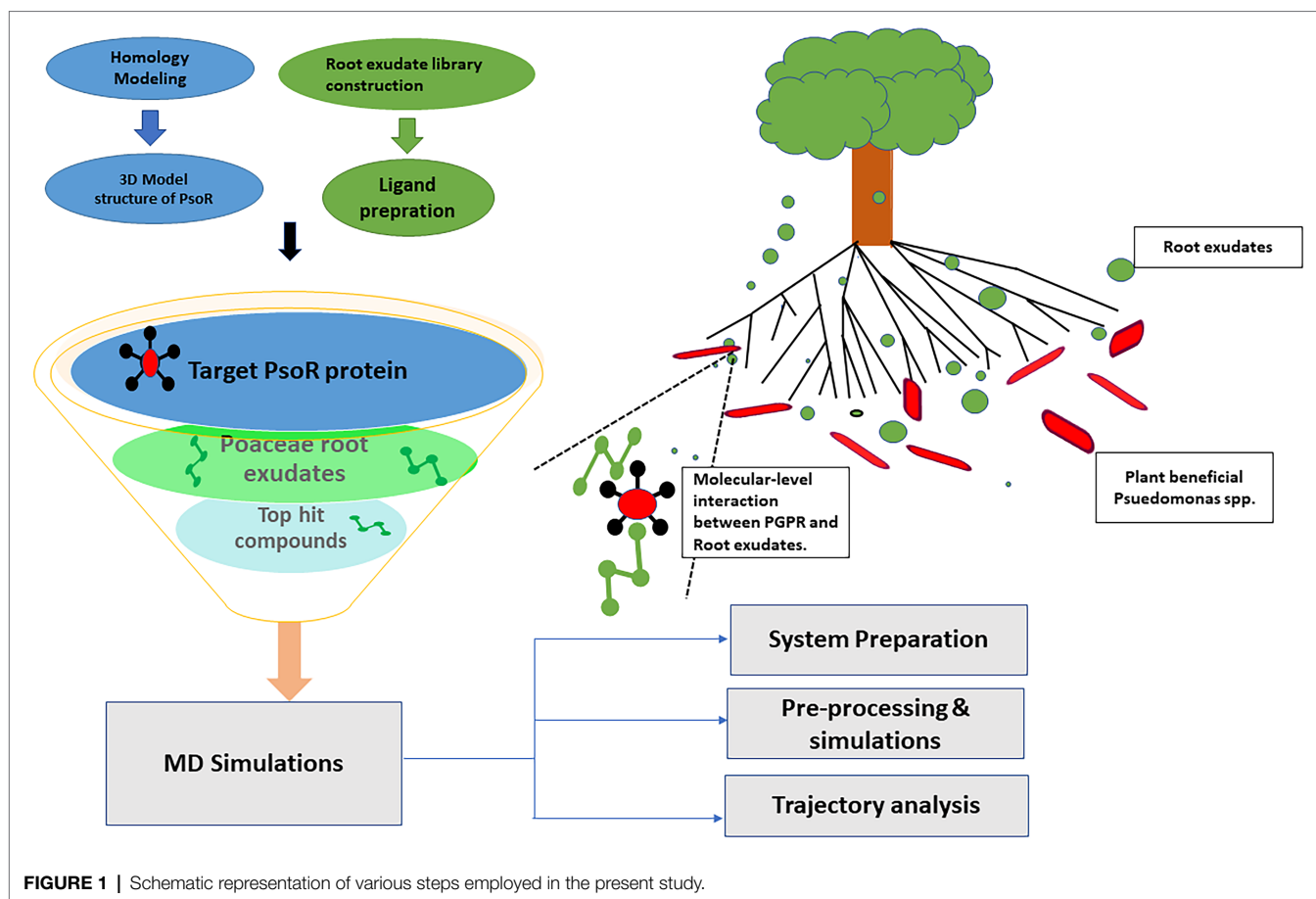
FIGURE 1 | Schematic representation of various steps employed in the present study.

MATERIALS AND METHODS

Molecular Modeling

Since the three-dimensional (3D) structure of PsoR of *Pseudomonas* is not available so far, we first procured the sequence of PsoR of *Pseudomonas protegens* Pf-5 from NCBI (accession number–WP_011063531.1), and using the "Easy Modeller" homology modeling tool, we later generated the 3D structure of the protein (**Figure 2A**; Kuntal et al., 2010). For the additional refinement of the artificially constructed model of PsoR, it was subjected to energy minimization, side-chain refinement, loop refinement, etc. To authenticate the refined structure, it was uploaded into the PDBsum server (Laskowski et al., 2018).

Active Site Analysis

For verifying the active sites within the modeled PsoR protein and for tracing, characterizing, and quantifying its spatial attributes, the CASTp webserver was employed (Lozano et al., 2012). Various catalytically active amino acid residues were identified with the help of CASTp 3.0, which might be involved in protein–ligand interaction. For rigorous docking, the estimated pocket attached with the amino acid residues linked with the reference molecule was taken into account.
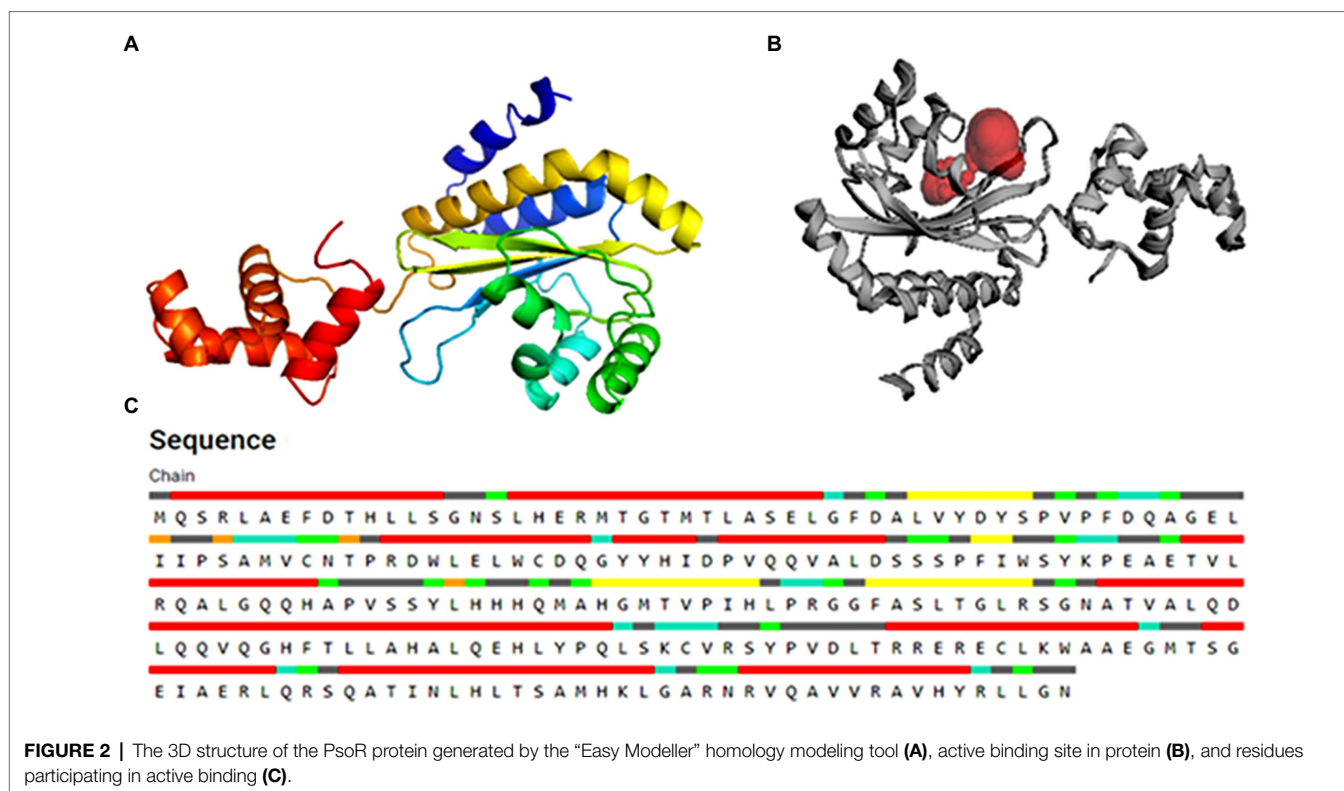
Ligand Preparation

For the selection of the low molecular, root exudates of Poaceae, we did an extensive literature survey on root exudates of barley, rice, rye, maize, and wheat plants. We created a phytochemical library of total of 59 diverse plant root exudates to be used in our study. The 3D structures of the root exudates of Poaceae and the reference molecule in the SDF format were retrieved from NCBI Pubchem.[1] Using Open Babel software (Version 2.3.1), all root exudates and the reference molecule in the SDF files were translated to PDB format.

Molecular Docking and Visualization

By employing PyRx open-source software, molecular docking was carried out to generate a set of potential ligand topologies and conformations at the binding site (GUI version 0.8 of AutodockVina; Trott and Olson, 2010). The calculation of the bound configurations by the PyRx software relies upon the binding affinity. For molecular docking studies, the grid center was set to $X = 10.84$, $Y = -24.12$, and $Z = 13.30$, and the dimensions of the grid box were set as 25.00, 25.30, and 25.96 Å with a spacing of 0.375 Å between the grids points. The number of exhaustiveness was set to eight by default. The polar hydrogen

[1]https://pubchem.ncbi.nlm.nih.gov

FIGURE 2 | The 3D structure of the PsoR protein generated by the "Easy Modeller" homology modeling tool **(A)**, active binding site in protein **(B)**, and residues participating in active binding **(C)**.

atoms and Kollamen charges were assigned during receptor and ligand preparation. Compounds were computationally screened with rigorous molecular docking in the catalytic center of the PsoR protein while allowing ligand molecules to be flexible. Total nine bound conformations of each ligand were generated using this program based on binding affinity. For further investigation, only the lowest binding energy conformational states of the ligand molecules were chosen. The Ligplotv.1.4.5 program was used to visualize and examine the molecular characteristics among protein–ligand complexes, such as hydrogen bonds and bond lengths (Karplus and McCammon, 2002). However, the 3-D protein–ligand complex was studied using PyMol molecular visualization tool (Pronk et al., 2013).

Molecular Dynamics Simulations

Molecular dynamics simulations were used to determine the structural stability of the protein–ligand complex under varying physiological circumstances (Karplus and McCammon, 2002). MD simulations were used to test the top compounds found by molecular docking. GROMACS was employed to conduct all of the MD simulations (Jorgensen et al., 1983). To generate topologies for the protein–ligand complex, the CHARMM 36 force field was used (Kumari et al., 2014). In the topology of the protein–ligand complex, the entire bonded and nonbonded factors were described, e.g., valency, atom groups, and bond connectivity. To mimic the biological environment, molecules have been dissolved in water. The water model of TIP3P was then used to construct a water solvated system with dodecahedral periodic boundary conditions and box vectors of equal length

9.81 nm. The solutes were placed in the simulation box with a minimum distance to the box edge of 10 Å (1.0 nm). The addition of three Na^+ ions neutralized each solvated system. After adding the ions, we undertook energy minimization on the protein–ligand complex to confirm that the scaffold had no steric clashes and an appropriate starting structure. This stage was processed at 10 kJ/mol using the steepest descent algorithm and the Verlet cutoff method. Furthermore, the equilibration of the protein–ligand complex was accomplished in two stages. Equilibration was carried out at 300 K for 100 ps under the NVT ensemble. It kept the system's temperature in balance. The NPT ensemble used Parrinello–Rahman simulation to perform second-phase equilibration. Here also the system was treated with a constant temperature (300 K) and pressure (1 atm) with a 2 fs time phase. The MD simulations of the protein–ligand were carried out for 100 ns. The root-mean-square divergence (RMSD), the radius of gyration (RG), root-mean-square fluctuation (RMSF), hydrogen bonds, solvent accessible surface area (SASA), and principal component analysis (PCA) were quantified to analyze the stability of the enzyme and protein–ligand complex. With the help of hydrogen bond analysis, the total number of unique hydrogen bonds generated during the protein–ligand reaction in the MD simulation was estimated.

Binding Free Energy Calculation Using MM-PBSA

The MMPBSA (molecular dynamics Poisson–Boltzmann surface area) method is commonly used to determine the binding free energy and predict the stability of the protein–ligand

complex following MD simulation (Kumari et al., 2014). Following that, using the molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) procedure provided in the g_mmpbsa package, an extensive analysis of the binding free energy (DGbind) was done (Kuzmanic and Zagrovic, 2010). For the calculation of binding free energy of protein–ligand complexes, both free solvation energy (polar and nonpolar solvation energies) and potential energy (electrostatic and Van der Waals interactions) are required, which were measured by the MMPBSA approach. We evaluated the nonbonded interaction energy between protein and ligands after MD simulation was completed, using conditions analogous to MD simulation for computing the extent of the interaction between protein complexes.

RESULTS

In the current study, we docked the modeled PsoR protein with root exudates of Poaceae and investigated the interactions undertaking between PsoR-saponarin and PsoR-BOA at the molecular level using MD simulations.

Homology Modeling

The primary structure of the PsoR protein of *P. protegens* Pf-5 was retrieved from NCBI, and BLASTP was used to recognize its cognate templates from the Protein Data Bank (PDB) database. The PsoR query is comparable to the crystal structure of PsoR in plant-beneficial *Pseudomonas* spp., according to BLASTP findings (PDB ID 6MWZ, 3SZT, and 4Y17), with 29.65, 29.08, and 27.15% identity and 88, 76, and 84% query coverage, consecutively. PDB ID 6MWZ corresponds to QS receptor protein, LasR of *Pseudomonas aeruginosa* UCBPP-PA14. Similarly, PDB ID 3SZT corresponds to QS control repressor, QscR of *P. aeruginosa* ATCC 15692, whereas PDB ID4Y17 is of a QS transcriptional regulator, SdiA, found in *Escherichia coli* K12. Later, Easy Modeller software was put to use for predicting the 3D structure of PsoR by considering these three structures as templates. Additionally, the Swiss-Pdb Viewer program (Guex and Peitsch, 1997) was used to minimize the energy of the predicted structure. The accurateness of the predicted structure is then authenticated using the ProQ (Wallner and Elofsson, 2003) and ProSA-web (Wiederstein and Sippl, 2007) servers. The quality of the projected model was analyzed using LG and MaxSub scores, which were estimated to be about 9.991 and −0.605, respectively, placing the model in the good prediction area. Likewise, the ProSA-web server presents the Z-score as a criterion to determine the total value of the PsoR protein, having a magnitude of −7.56. This suggests that the 3D structure has been accurately predicted. The 3D model of protein also demonstrated that about 92.8% of the total residues fall within the permitted region of the Ramachandran plot (**Figure 3**), further confirming the stereochemical stability of the PsoR homology template.

Active Site Confirmation

The substrate accessible pockets and catalytic sites on the putative PsoR protein were analyzed using the Computed Atlas of Surface Topography of Proteins (CASTp) algorithm. Following are the amino acid residues in the pocket, which were selected for virtual screening: Val38, Asp40, Ser42, Pro43, Ala49, Gly50, Glu51, Leu52, Ile54, Leu67, Trp70, Tyr76, His77, Asp79, Val81, Gln82, Ala85, Leu86, His120, His121, Thr129, Phr139, Ser141, and Ser143 (**Figure 2C**). PsoR has an active site area of 184.972 and a volume of 118.953, according to CASTp findings (**Figure 2B**). 4-[3-(methylsulfonyl)phenoxy]-N-[(1S,3S,5S)-2-oxobicyclo[3.1.0]hexan-3-yl]butanamide (K5M) was taken as a reference compound as it was co-crystallized with PDB ID-6MWZ because this protein showed highest query cover with PsoR protein.

Molecular Docking

About 59 substrates were enriched by docking with the PsoR receptor (**Supplementary Table 1**). Among the total 59 root exudates, the best five were chosen from molecular docking, having better or nearly equal binding energy in comparison to the reference molecule K5M (−7.0 kcal mol⁻¹; **Table 1**), and only those five exudates were further subjected to MD simulations. The MD simulations revealed that out of five, only two root exudates, i.e., saponarin and 2-benzoxazolinone (BOA), showed suitable binding with PsoR protein. Hence, we have targeted only these two ligands, i.e., saponarin and 2-benzoxazolinone (BOA) in a later study (**Figure 4**). The molecular docking study of PsoR with saponarin exposed that the catalytic site of the PsoR protein is made up of 5 amino acid residues, *viz.* Pro43, Phe46, Ala49, His77, and Gln82 amino acids, which are stabilized by five hydrogen bonds, whereas Ser42, Val44, Asp47, Gln48, Leu52, Tyr76, Gln83, Leu86, and Phe139 take part in hydrophobic interaction. Likewise, BOA exhibited association with Ser141 by one H-bond, although Val38, Asp40, Tyr76, Asp79, Val81, Gln82, His121, Thr129, and Thr143 residues were linked to PsoR through hydrophobic interaction (**Figure 5**). The results imply that the tested compounds may act similarly to the reference molecule K5M.

Molecular Dynamic Simulation

Molecular dynamics simulation was used to assess the stability of the PsoR-exudate complex. It was used to study the physical mobility of atoms and molecules, as well as to predict conformational changes at the molecular level. As the binding model of BOA and saponarin was shown to be suitable based on PyRx's molecular docking performance, the structural alterations of only these two exudates were further examined in this study. The binding stability and free energy of protein–ligand complexes were determined using a molecular dynamics simulation with a time step of 100 ns. The protein–ligand complex in PsoR activation and its effects in the simulation was studied using RMSD, RMSF, RG, SASA, and MMPBSA measurements. As already mentioned, according to the MD simulation data, only two exudates were found to be stable against the PsoR enzyme (**Figure 4**). Consequently, in the following study, only the effects of those two exudates, saponarin and BOA, were reported.
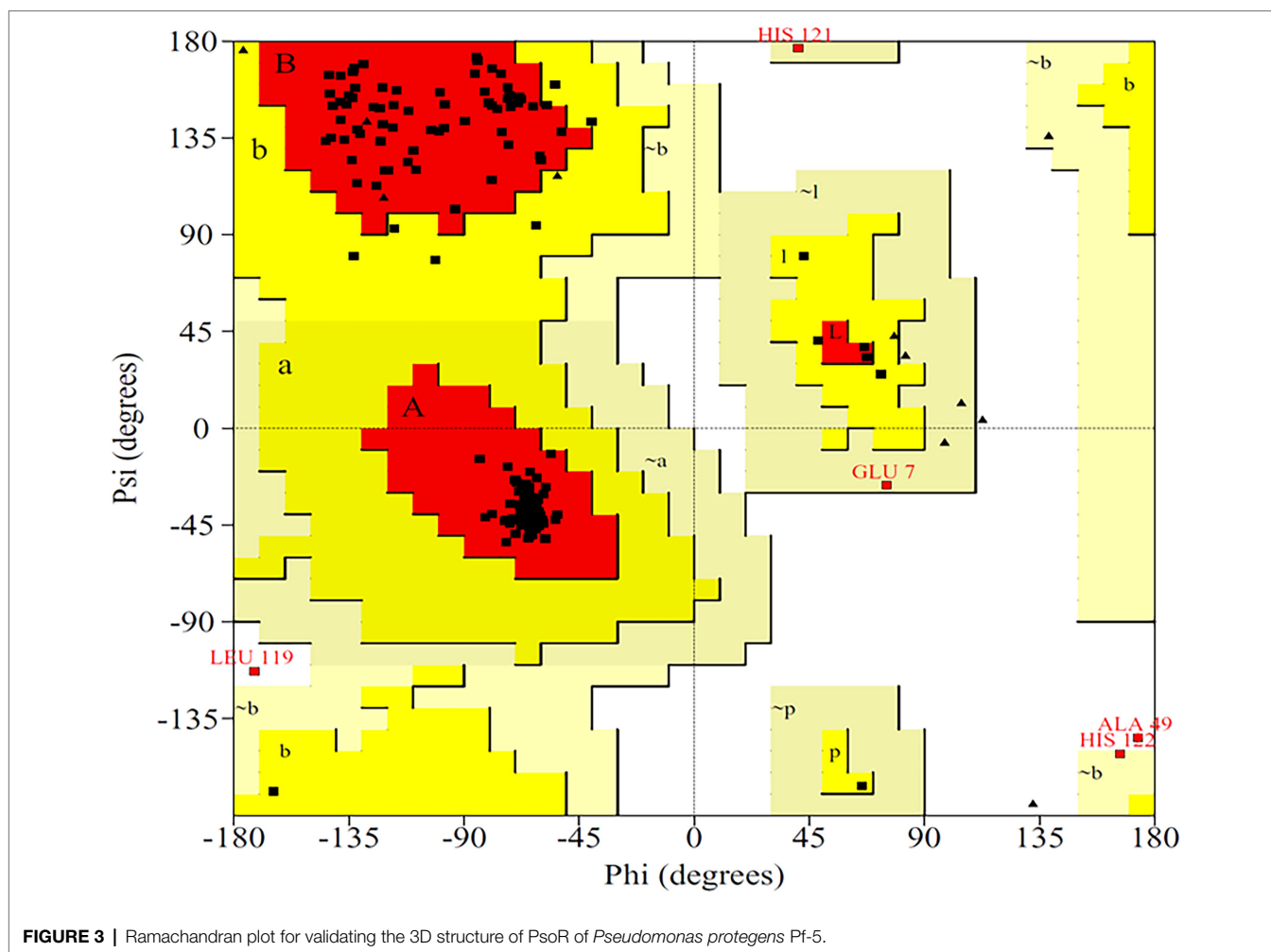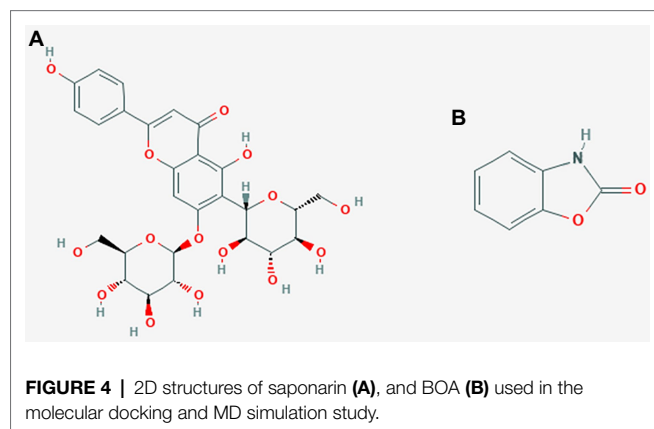
FIGURE 3 | Ramachandran plot for validating the 3D structure of PsoR of *Pseudomonas protegens* Pf-5.

TABLE 1 | Molecular docking scores of various screened root exudates with PsoR of *Pseudomonas protegens* Pf-5.

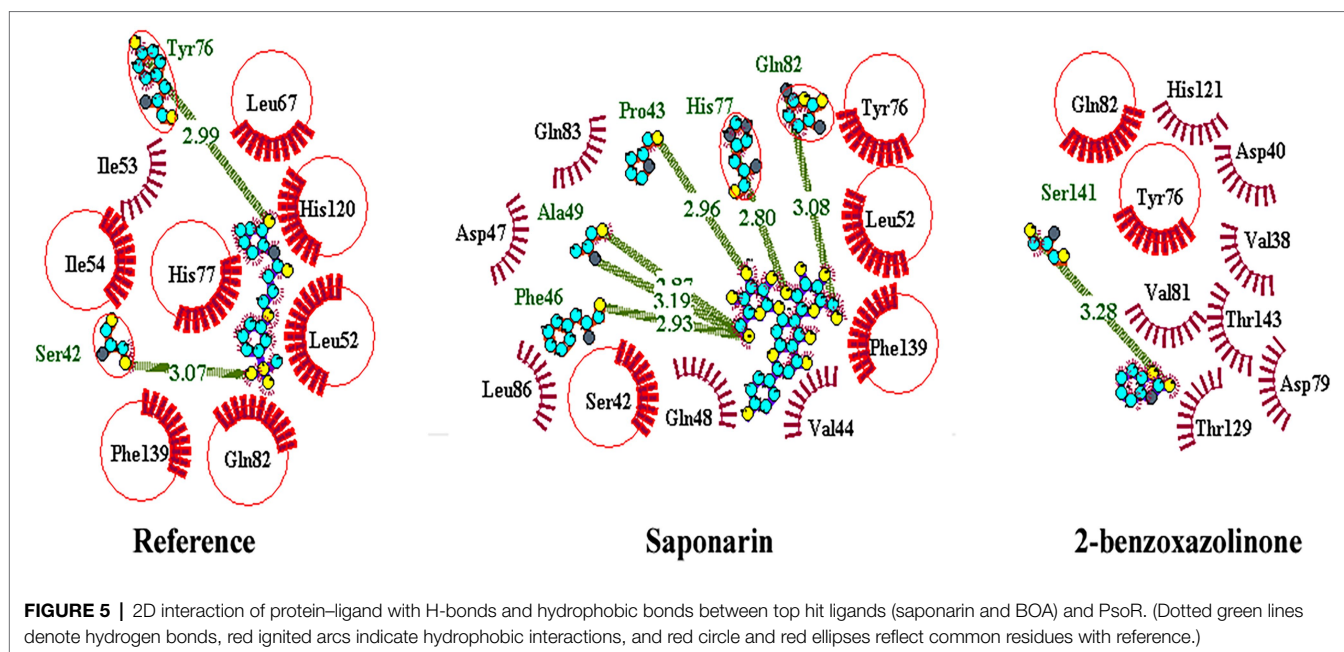| S. no.	Name of the substrate	Compound ID	Molecular formula	Binding energy, kcal mol^{-1}
1	Reference (K5M)	137,628,326	$C_{17}H_{21}NO_5S$	−7.0
2	Saponarin	441,381	$C_{27}H_{30}O_{15}$	−7.1
3	o-coumaric acid	637,540	$C_9H_8O_3$	−6.6
4	2-Benzoxazolinone	6,043	$C_7H_5NO_2$	−6.4
5	trans-cinnamic acid	444,539	$C_9H_8O_2$	−6.4
6	p-hydroxybenzoic acid	135	$C_7H_6O_3$	−6.3
7	Adenosine	60,961	$C_{10}H_{13}N_5O_4$	−6.2
8	Phthalic acid	1,017	$C_8H_6O_4$	−5.8
9	Ferulic acid	445,858	$C_{10}H_{10}O_4$	−5.6
10	D-maltose	6,255	$C_{12}H_{22}O_{11}$	−5.5

FIGURE 4 | 2D structures of saponarin **(A)**, and BOA **(B)** used in the molecular docking and MD simulation study.

RMSD of the Protein–Ligand Complex

The variation in the protein–ligand complex was assessed using the root-mean-square deviation (RMSD) at 100 ns. The variations generated in a protein during simulation characterize its structural stability. The nature of proteins with fewer variations is more predictable. The RMSD value was computed

for the α-carbon of saponarin and BOA with PsoR. **Figure 6A** shows the curve of RMSD (nm) vs. time (ps) for both ligands. The RMSD trajectory of the protein and all of the complexes attained equilibrium, as shown in the diagram. For the complexes, PsoR-BOA and PsoR-saponarin, the average RMSD values were 0.59 ± 0.09 nm, and 0.65 ± 0.13 nm, respectively,

FIGURE 5 | 2D interaction of protein–ligand with H-bonds and hydrophobic bonds between top hit ligands (saponarin and BOA) and PsoR. (Dotted green lines denote hydrogen bonds, red ignited arcs indicate hydrophobic interactions, and red circle and red ellipses reflect common residues with reference.)

(**Table 2**). RMSD is a measure in MD simulations to evaluate the equilibration and mobility of proteins/enzymes and even to estimate the distance between the protein's backbone and atoms (Sargsyan et al., 2017). Both the protein–ligand complexes, namely PsoR-BOA and PsoR-saponarin, are stable, as evidenced by a small variation and a reduced RMSD. Furthermore, the results confirmed that the MD simulations for PsoR remained constant at 100 ns during the reaction which indicates that the PsoR enzyme can interact with saponarin and BOA ligands. To better understand the variation in the complex during biological processes, other investigators have also evaluated the RMSD for protein–ligand complexes (Joshi et al., 2021).

RMSF of the Protein–Ligand Complex

To predict the movement of an atom's position at a specific temperature and pressure for two sets of PsoR-exudate complexes, an RMSF evaluation was done. The RMSF findings depicted the flexible areas of the protein and estimated the overall variation of protein during the MD simulation. A lower RMSF reading is often considered good as it signifies that the protein–ligand combination was more stable, while a larger reading means MD was more flexible. For the protein–ligand complexes, variations in the component residue of the enzyme were observed over the 100 ns trajectory time (**Figure 6B**). For the duration of the encounter of the PsoR-exudate complex, a variation of less than 0.4–0.2 nm was detected in the active binding region, which is perfectly acceptable (**Table 3**). Therefore, as per the RMSF readings, it can be said that both the complexes were stable throughout MD simulations, with reduced fluctuation and increased stability. RMSF has been used to investigate protein–ligand complexes in the past. The findings demonstrate that when the flexibility of an enzyme is hindered, conformational rearrangements for substrate binding occur during catalysis.

Radius of Gyration

The radius of gyration indicates the degree of compactness of an enzyme–substrate complex. It deals with the folding and unfolding of proteins. The radius of gyration was estimated by utilizing the final 100 ns trajectories in this study. The mean RG readings for BOA and saponarin were 1.8 ± 0.004 and 1.8 ± 0.006 nm, correspondingly (**Table 2**). The steadiness in the RG reading indicates folding of the enzyme, whereas fluctuations in the RG reading correspond to the unfolding of enzymes (Lobanov et al., 2008). The RG results uncovered that every protein–ligand complex displayed an RG value that was comparatively close and in line with the expected statistics (**Figure 6C**). This indicates that both the exudate compounds were completely superimposed over PsoR protein which reflected in compaction and stabilization of the protein–ligand complex (Pande et al., 2021).

Solvent Accessible Surface Area

Solvent accessible surface area (SASA) is the abbreviation for solvent accessible surface area. For the duration of MD simulation, the total surface area of protein that can be admitted by the water solvent is evaluated by SASA and it can be used further for investigating communication within the solvent and the complex. For BOA, the mean SASA was found to be 142.15 ± 3.7 nm^2, whereas for saponarin it was evaluated as 152.62 ± 2.7 nm^2 (**Table 2**). Hence, in our findings the SASA values for the two aforementioned complexes remained astonishingly stable in the 100 ns MD simulation, implying that the structure of the protein has not changed considerably (**Figure 6D**). The PsoR complexes carried an elevated SASA reading in both cases. This may be due to the gradual unfolding of the critical amino acids in the active sites of PsoR protein, after interacting with BOA and saponarin. As a result, polar
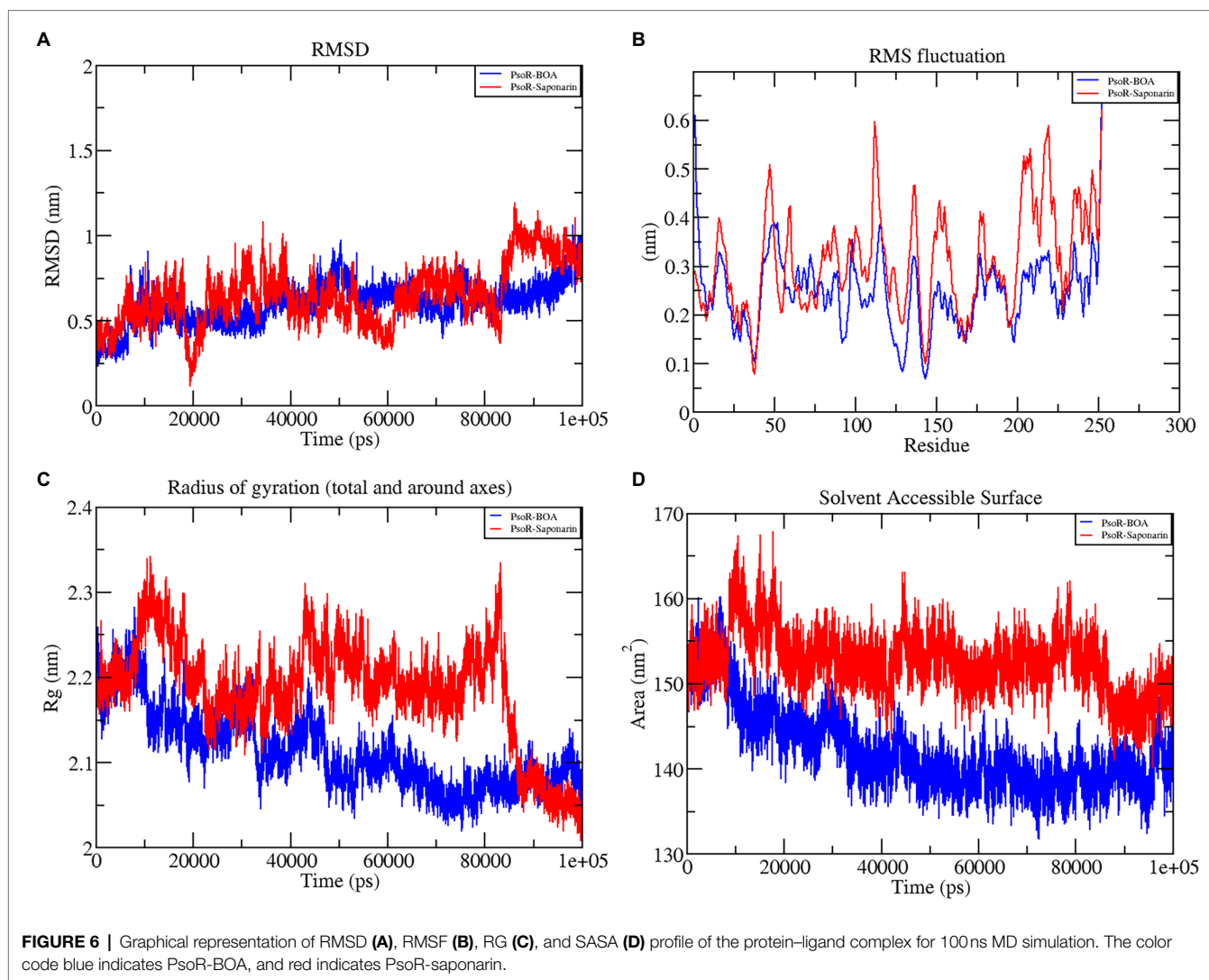
FIGURE 6 | Graphical representation of RMSD **(A)**, RMSF **(B)**, RG **(C)**, and SASA **(D)** profile of the protein–ligand complex for 100 ns MD simulation. The color code blue indicates PsoR-BOA, and red indicates PsoR-saponarin.

TABLE 2 | The average values of RMSD, RG, SASA, and interaction energy of different protein–ligand complexes.

Protein–ligand complex	Average RMSD (nm)	Average RG (nm)	Average SASA (nm²)	Interaction energy (kJ/mol)
PsoR-2-benzoxazolinone	0.59 ± 0.09	1.8 ± 0.004	142.15 ± 3.7	−71.1813
PsoR-saponarin	0.65 ± 0.13	1.8 ± 0.006	152.62 ± 2.7	−167.032

water molecules get entry into the newly opened pocket. This indicates that the channels and cavities of the proteins are essential in recognition of bacterial PsoR protein and its further binding with plant root exudates like BOA and saponarin. In conclusion, we have observed that PsoR's surface has generated cavities.

The results of this study demonstrate that as a result of GROMACS' charm force field electrostatic interaction, PsoR complexes develop spontaneously in an aqueous solution. These findings shed light on the interaction of BOA and saponarin

within the enzyme cavity, and also on the electrostatic attraction aiding in stable binding.

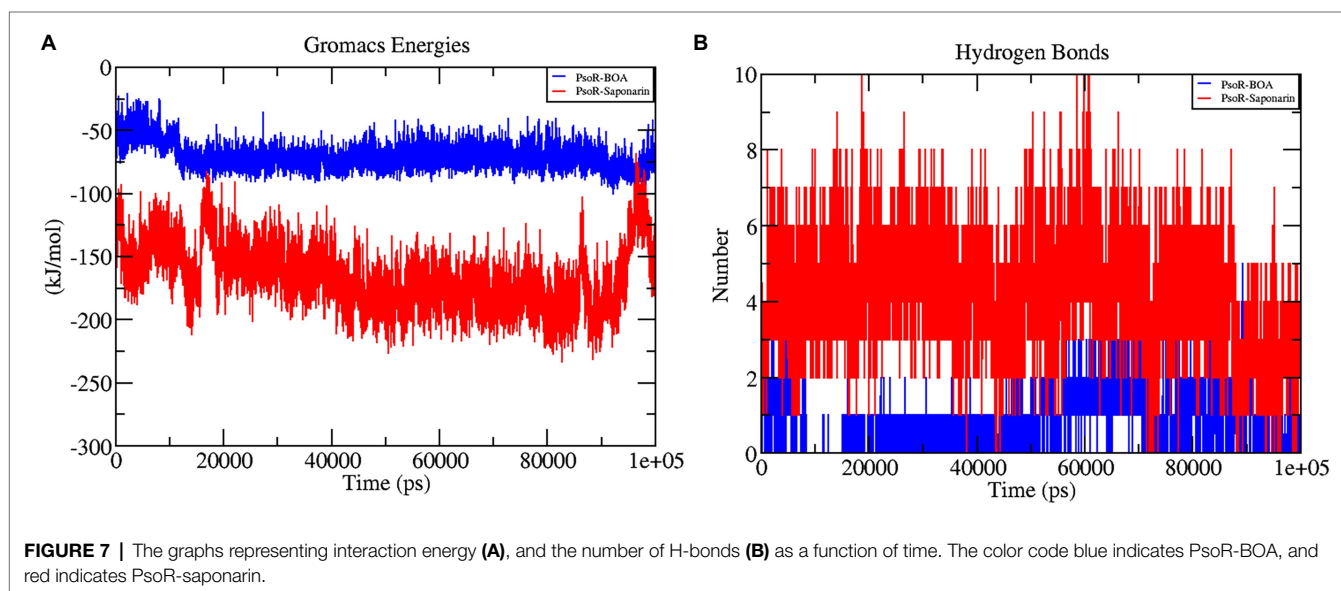## Interaction Energy of Protein–Ligand Complex
The interaction energy is used to analyze the efficacy of the protein–ligand combination. For confirmation of the strength of protein and ligand molecule, interaction energy was used. The interaction energy was calculated using GROMACS' Parrinello–Rahman parameter. The estimated interaction energy values for the BOA and saponarin were $-71.1813\,\mathrm{kJ\,mol^{-1}}$ and $-167.032\,\mathrm{kJ\,mol^{-1}}$, respectively, in a 100 ns simulation session (**Figure 7A**; **Table 2**). The interaction energy results suggest that PsoR efficiently recognizes and binds with BOA and saponarin. As a consequence of these findings, an inter-kingdom signaling circuit between root exudates and plant-associated bacteria can be further explored.

Hydrogen Bonding Analysis of Protein–Ligand Complex
In the interaction of enzyme and substrate, hydrogen bonding is considered critical (Chen et al., 2016). It is required for

TABLE 3 | Root-mean-square fluctuation (RMSF) values (nm) of residues involved in the interaction of the protein–ligand complex.

Protein–ligand complex	Hydrogen bond interaction			Hydrophobic bond interaction		
	No. of bond	Residues involved	RMSF value	No. of bond	Residues involved	RMSF value
PsoR-2-benzoxazolinone	1	Ser141	0.11	9	Val38	0.10
					Asp40	0.15
					Tyr76	0.23
					Asp79	0.26
					Val81	0.20
					Gln82	0.21
					His121	0.18
					Thr129	0.08
					Thr143	0.06
PsoR-saponarin	5	Pro43	0.33	9	Ser42	0.27
		Phe46	0.46		Val44	0.40
		Ala49	0.40		Asp47	0.50
		His77	0.24		Gln48	0.47
		Gln82	0.32		Leu52	0.25
					Tyr76	0.21
					Gln83	0.35
					Leu86	0.36
					Phe139	0.29

FIGURE 7 | The graphs representing interaction energy **(A)**, and the number of H-bonds **(B)** as a function of time. The color code blue indicates PsoR-BOA, and red indicates PsoR-saponarin.

the substrate's selectivity, metabolism, and catalysis. The two complexes were explored further for a better understanding of the hydrogen bond in their association with PsoR protein. The total H bonds were recorded at the end of 100 ns. Saponarin contains 10 H bonds, whereas BOA has four H bonds, as per the findings (**Figure 7B**). These results indicate that saponarin binds to the PsoR active site more efficiently than BOA.

Principal Component Analysis

The principal component analysis (PCA) study helps in examining the important coordinated movements that occur during ligand binding. Only the first few eigenvectors are known to define the overall motion of the protein. In this study, the eigenvectors are obtained by diagonalizing the matrix. This study used the top 40 eigenvectors to estimate coordinated movements. **Figure 8A** shows the eigenvalues generated from the diagonalization of the covariance matrix of atomic fluctuations versus the appropriate eigenvector in decreasing order. Out of the 40 eigenvectors, the first 10 accounted for 88.15 and 91.16% of overall motions for BOA and saponarin, respectively. From PCA findings, this can be safely inferred that both of the studied complexes had lesser movements and form a stable association with PsoR. This could only be possible if the ligand association has altered protein structure and dynamics. Using PCA to construct 2D projection plots is another approach for obtaining the dynamics of protein–ligand complexes. **Figure 8B**

FIGURE 8 | Principal component analysis: plot of the eigenvalues vs. first 40 eigenvectors **(A)**, the first two eigenvectors showing the protein motion in phase space for all the complexes **(B)**. The color code blue indicates PsoR-BOA, and red indicates PsoR-saponarin.
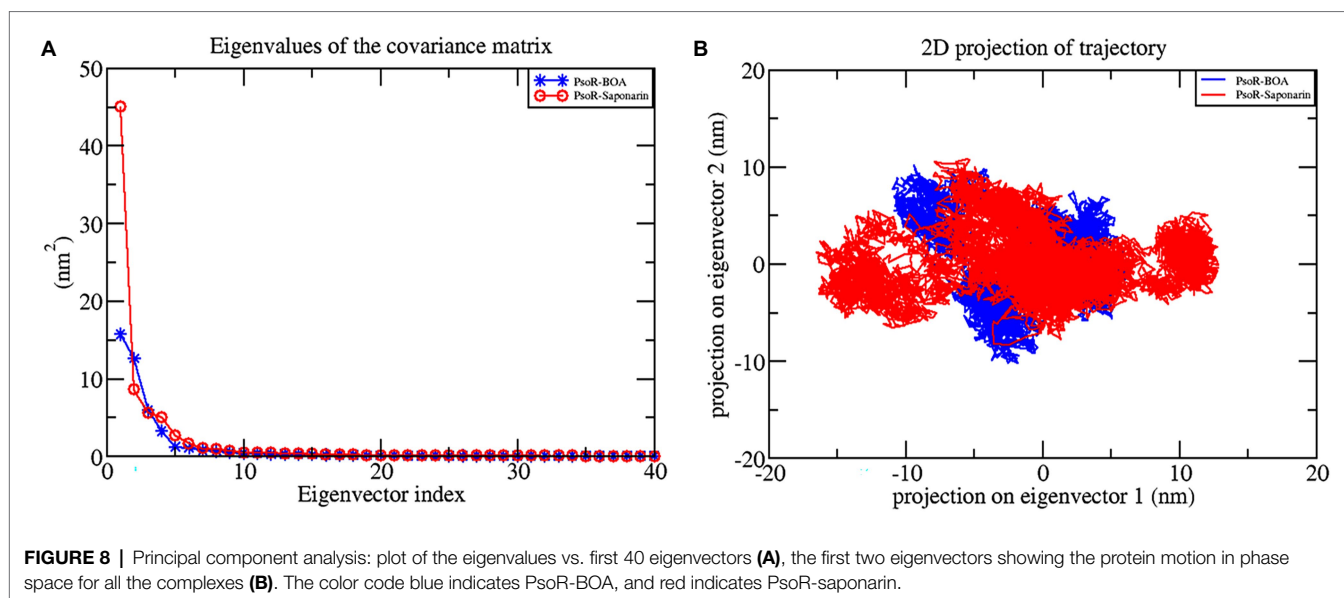
depicts a two-dimensional projection of the trajectories in phase space for the first two main components, PC1, and PC2, for both complexes. In general, the complexes occupying a compact phase space along with a stable cluster are considered more stable, whereas those complexes taking greater space with unstable clusters are considered less stable. As they took up minimal area in the phase space, all of the complexes were found to be very stable. The above PCA and other MD simulation findings were also in accordance with the 2D PCA reports. **Figure 9** shows the Gibbs energy landscape plot for PC1 and PC2 degrading complexes for saponarin and BOA. The results show that PsoR-saponarin and PsoR-BOA have a Gibbs free energy of 0–13.6 kJ mol^{-1} (**Figure 9A**) and 0–13.5 kJ mol^{-1} (**Figure 9B**), respectively. The free energy profiles of both protein–ligand complexes were significant, showing that the complexes remained stable at the time of the catalysis of saponarin and BOA. According to our studies, the PsoR complex was found to be thermodynamically efficient.

Binding Free Energy

The binding free energy was estimated from MD trajectories with the help of the MM-PBSA method executed in GROMACS. The overall binding energy of all protein–ligand complexes was found to be in the acceptable range. Precisely, complex PsoR-BOA showed the least binding free energy and greatest binding affinity with PsoR (−31.600 kJ mol^{-1}), signifying a well-stabilized ligand conformation. For saponarin, the binding free energy was found to be −32.41 KJ mol^{-1} (**Table 4**). The results of the molecular docking and MD simulations were verified by the binding free energy calculation, confirming that these compounds predominantly bind to the PsoR. The MM-PBSA method has been utilized to quantify the exact binding free energy in biological reaction analysis. For the MM-PBSA study, the PsoR-saponarin and PsoR-BOA complexes were utilized, and the findings confirmed the predominant role played by PsoR protein in recognizing and binding further to at least two root exudates from the Poaceae

family. We conclude that PsoR-saponarin and PsoR-BOA are stable complexes with remarkable binding affinities based on the overall MD Simulation (containing RMSD, RMSF, and RG analysis), post-MD simulation (including hydrogen bonds, SASA, gap, and PCA analysis), and binding free energy analysis findings.

DISCUSSION

Shortly after the discovery of LuxR solo proteins (LuxR orphan protein), it was realized that it interacts with some low molecular weight compounds from root exudates of the host plant and thus enables chemical communication between them, opening a new avenue toward the inter-kingdom signaling (González et al., 2013). PsoR is a LuxR solo regulator majorly present in plant-beneficial *Pseudomonas* spp. Interestingly, the presence of multiple PsoR orthologs in the genetic pool of plant-associated bacteria signifies both its widespread nature and possible role in mediating plant-bacterial signaling (Subramoni and Venturi, 2009). The studies on PsoR action have been limited as the crystal 3D structure was not available in the PDB databank. In this context, we have built a 3D model of PsoR for the first time, by using homology modeling. The modeled protein was subjected to energy minimization using the Swiss-PDB Viewer program. Further, to authenticate the accurateness of the predicted structure its ProQ and ProSA-web servers were employed. ProSA estimated Z-score value, −7.56 evincing highly reliable structure. ProQ generated LG and MaxSub scores of 9.991 and −0.605.

The Ramachandran plot created by the PDBsum server shows details on the arrangement of amino acid residues in the backbone dihedral angles Phi against Psi in the protein structure. Ramachandran dihedral statistics for modeled 3D structure of PsoR revealed a total of 92.8% residues in permitted regions. The results of the RC plot indicated that the built model had great geometry and was precise and reliable for
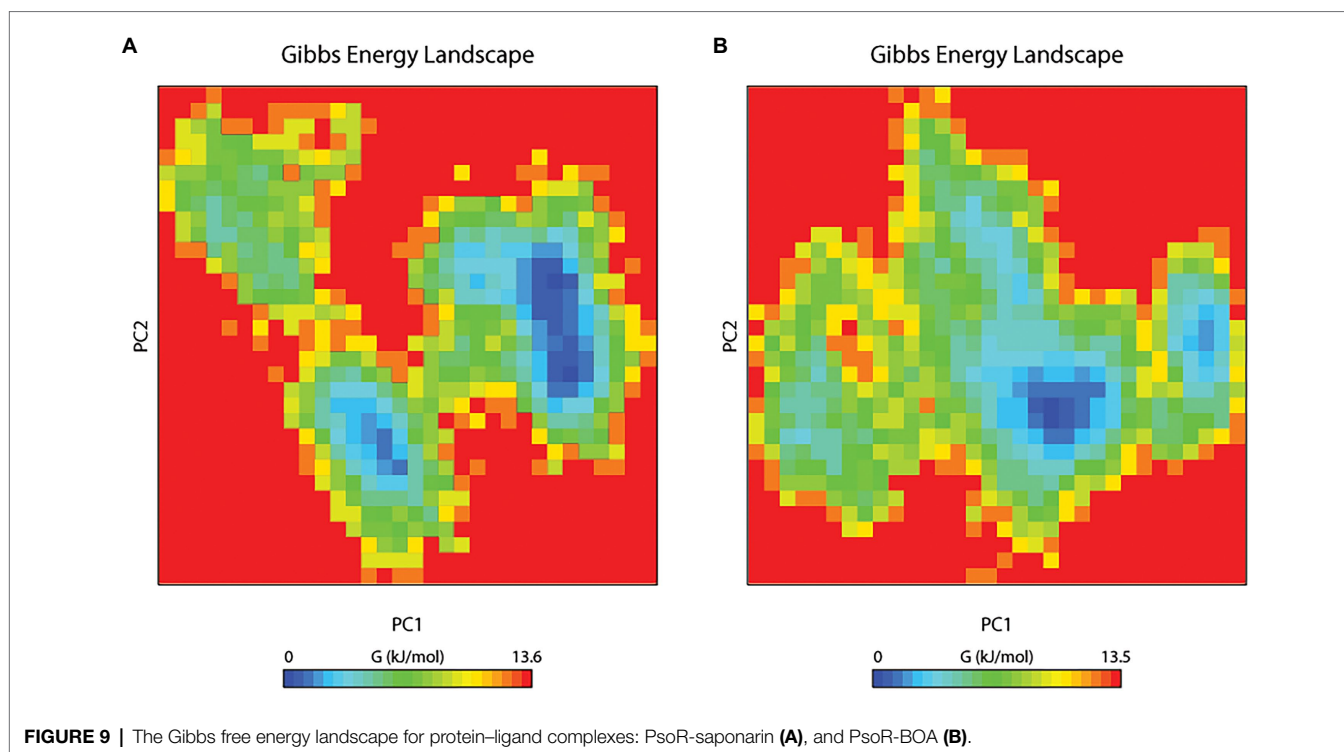
FIGURE 9 | The Gibbs free energy landscape for protein–ligand complexes: PsoR-saponarin **(A)**, and PsoR-BOA **(B)**.

TABLE 4 | Table representing the Van der Waals, electrostatic, polar salvation, SASA, and binding energy for protein–ligand complexes.

Protein–ligand complex	Van der Waals energy (kJ mol^{-1})	Electrostatic energy (kJ mol^{-1})	Polar solvation energy (kJ mol^{-1})	SASA energy (kJ mol^{-1})	Binding energy (kJ mol^{-1})
PsoR-2-benzoxazolinone	−84.995 ± 7.863	−14.054 ± 13.469	77.036 ± 17.254	−9.586 ± 0.558	−31.600 ± 8.779
PsoR-saponarin	−141.523 ± 21.995	−61.966 ± 31.854	191.573 ± 44.689	−20.503 ± 2.115	−32.418 ± 19.219

future docking and MD simulations. Earlier studies suggested that the PsoR is likely specific for a molecule present in Poaceae plants. Therefore, to uncover a possible PsoR elicitor, we conducted a literature search and created a phytochemical library comprising 59 root exudates from barley, rice, rye, maize, and wheat plants. The molecular docking of the modeled PsoR, from AutoDock Vina, resulted in the best five compounds with lower or comparable binding energy to the reference molecule. As only two root exudates, i.e., saponarin and 2-benzoxazolinone (BOA), showed suitable binding with PsoR protein in MD simulations, we have elaborately discussed only these two exudates, i.e., saponarin and 2-benzoxazolinone (BOA) in our study. According to the LigPlot data, both these exudates can establish hydrogen bonds on the active site of the protein, implying that they could activate the enzyme. The reference molecule formed two hydrogen bonds with Ser42 and Tyr76. In our results, saponarin formed five hydrogen bonds with Pro43, Phe46, Ala49, His77, and Gln82, whereas BOA formed only one hydrogen bond with Ser141 and predominantly made hydrophobic interactions with Val38, Asp40, Tyr76, Asp79, Val81, Gln82, His121, Thr129, and Thr143. Both exudates displayed negative binding energy, making hydrogen and hydrophobic associations with active sites, suggesting a better

association with PsoR. The more negative binding energy implied a relatively stable protein–ligand combination. In this study, both the complexes had fewer variations and lower RMSD during 100 ns, pointing that the PsoR enzyme can interact with both the exudates. During the interaction of PsoR with saponarin and BOA, the overall fluctuation recorded was between 0.4 and 0.2 as per the RMSD value; therefore, both exudates remained stable throughout MD simulations. The RG value of both exudates remained steady and got completely superimposed over PsoR protein. The increase in SASA value suggests that after encounter of PsoR with BOA and saponarin has led to a gradual unfolding of the critical residues in the active sites of PsoR protein. This contributes to stable binding of the small molecules to the protein. The PCA analysis identified the top 40 eigenvectors, with the first 10 constituting 88.15% of overall BOA movements, and 91.16% of total saponarin movements, which is only conceivable if the ligand interaction has altered protein structure and kinetics, suggesting that binding of saponarin and BOA, caused changes in the structure of PsoR.

Other recent studies have also utilized PCA to investigate protein–ligand complexes and found that every protein–ligand complex corresponds to a variable value of Gibbs free energy

(David and Jacobs, 2014). In conclusion, MD simulation of the related protein–ligand complexes might help characterize the interactions of many additional exudate chemicals with PGPR proteins. Molecular dynamics study concludes that these two compounds are potent novel small molecules that might be used as lead molecules for the elicitation of PsoR protein to manifest its role in biocontrol by establishing inter-kingdom signaling. Saponarin is a natural diglycoside flavone (apigenin6-C-glucosyl-7-O-glucoside) compound present in the barley leaves. Similarly, 2-benzoxazolinone (BOA) is the degradation product of DIBOA (2,4-dihydroxy-2H-1,4-benzoxazin-3(4H)-one), one of the most abundant benzoxazinoids (BXs; Fomsgaard et al., 2006). It is produced in the growing root and shoots tissues; where after the addition of sugar moieties it is either retained in vacuoles or released from the roots. Saponarin is the major flavonoid found in barley, while BOA is the major secondary metabolite of wheat, maize, and rye. Earlier, González et al. (2013) concluded that PsoR protein most probably binds with root exudates compounds from the Poaceae family. Our findings are in line with this study, as both our root exudates compounds fall under the family Poaceae (González et al., 2013). Likewise, Neal et al. (2012) concluded that BXs are especially important in the rhizoplane of maize plants, where they attract the plant-friendly bacterium *Pseudomonas putida*. Recently, it has also been speculated that BXs increase the plant's fitness by generating key metabolites, capable of inter-kingdom signaling (Schandry and Becker, 2020). In our study, we also found that 2-BOA binds effectively with the PsoR protein of *P. protegens* Pf5 and hence it is highly likely of showing cross-kingdom interaction. Based on the molecular docking and molecular dynamic results, we suggest that saponarin and BOA, the two root exudates of Poaceae, may have a potential eliciting effect in PsoR protein of plant-beneficial *Pseudomonas* spp.

CONCLUSION

In this study, we have first-ever predicted the molecular interaction between the PsoR of *P. protegens* Pf5 and the root exudates of barley and wheat. For future research, it will be significant to look for *in vitro* experimental data to further decipher the entire metabolic pathway triggering the activation of biocontrol effects of the PsoR protein of plant-beneficial *Pseudomonas* spp.

DATA AVAILABILITY STATEMENT

All relevant data is contained within the article. The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

DS was the one who came up with the idea, designed the experiment, and wrote the first manuscript. The experiments were carried out by TJ and SP, who also evaluated the data and wrote the manuscript. VP and SM assisted with the manuscript's writing and editing. MS and SC oversaw the research, assisted with data analysis, and wrote, reviewed, and edited the paper. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.875494/full#supplementary-material

REFERENCES

Chen, D., Oezguen, N., Urvil, P., Ferguson, C., Dann, S. M., and Savidge, T. C. (2016). Regulation of protein-ligand binding affinity by hydrogen bond pairing. *Sci. Adv.* 2:e1501240. doi: 10.1126/sciadv.1501240

Covaceuszach, S., Degrassi, G., Venturi, V., and Lamba, D. (2013). Structural insights into a novel interkingdom signaling circuit by cartography of the ligand-binding sites of the homologous quorum sensing LuxR-family. *Int. J. Mol. Sci.* 14, 20578–20596. doi: 10.3390/ijms141020578

David, C. C., and Jacobs, D. J. (2014). Principal component analysis: a method for determining the essential dynamics of proteins. *Methods Mol. Biol.* 1084, 193–226. doi: 10.1007/978-1-62703-658-0_11

Fomsgaard, I., Mortensen, A., Idinger, J., Coja, T., and Blumel, S. (2006). Transformation of benzoxazinones and derivatives and microbial activity in the test environment of soil ecotoxicological tests on *Poecilus cupreus* and *Folsomia candida*. *J. Agric. Food Chem.* 54, 1086–1092. doi: 10.1021/jf050914a

Gaete-Eastman, C., Morales-Quintana, L., Herrera, R., and Moya-León, M. A. (2015). In-silico analysis of the structure and binding site features of an α-expansin protein from mountain papaya fruit (VpEXPA2), through molecular modeling, docking, and dynamics simulation studies. *J. Mol. Model.* 21:115. doi: 10.1007/s00894-015-2656-7

González, J. F., Subramoni, S., and Venturi, V. (2013). An inter-kingdom signaling mechanism in rhizosphere pseudomonas. *Mol. Microb. Ecol. Rhizosphere* 1, 755–761. doi: 10.1002/9781118297674.ch71

Guex, N., and Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18, 2714–2723. doi: 10.1002/elps.1150181505

Haas, D., and Keel, C. (2003). Regulation of antibiotic production in root-colonizing *Peudomonas* spp. and relevance for biological control of plant disease. *Annu. Rev. Phytopathol.* 41, 117–153. doi: 10.1146/annurev.phyto.41.052002.095656

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi: 10.1063/1.445869

Joshi, T., Joshi, T., Sharma, P., Chandra, S., and Pande, V. (2021). Molecular docking and molecular dynamics simulation approach to screen natural compounds for inhibition of *Xanthomonas oryzae* pv. *Oryzae* by targeting

peptide deformylase. *J. Biomol. Struct. Dyn.* 39, 823–840. doi: 10.1080/07391102. 2020.1719200

Karplus, M., and McCammon, J. (2002). Molecular dynamics simulations of biomolecules. *Nat. Struct. Mol. Biol.* 9, 646–652. doi: 10.1038/nsb0902-646

Kumar, A., Kumar, S., Kumar, A., Sharma, N., Sharma, M., Singh, K. P., et al. (2018). Homology modeling, molecular docking and molecular dynamics based functional insights into rice urease bound to urea. *Proc. Natl. Acad. Sci. India Sect. B Biol. Sci.* 88, 1539–1548. doi: 10.1007/s40011-017-0898-0

Kumari, R., Kumar, R.Open Source Drug Discovery Consortium, and Lynn, A. (2014). g_mmpbsa- A GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.* 54, 1951–1962. doi: 10.1021/ci500020m

Kuntal, B. K., Aparoy, P., and Reddanna, P. (2010). EasyModeller: a graphical interface to MODELLER. *BMC. Res. Notes* 3, 1–5. doi: 10.1186/1756-0500-3-226/FIGURES/5

Kuzmanic, A., and Zagrovic, B. (2010). Determination of ensemble-average pairwise root mean-square deviation from experimental B-factors. *Biophys. J.* 98, 861–871. doi: 10.1016/j.bpj.2009.11.011

Laskowski, R. A., Jabłońska, J., Pravda, L., Vařeková, R. S., and Thornton, J. M. (2018). PDBsum: structural summaries of PDB entries. *Protein Sci.* 27, 129–134. doi: 10.1002/pro.3289

Lobanov, M. Y., Bogatyreva, N. S., and Galzitskaya, O. V. (2008). Radius of gyration as an indicator of protein structure compactness. *Mol. Biol.* 42, 623–628. doi: 10.1134/S0026893308040195

Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., et al. (2012). Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010. *Lancet* 380, 2095–2128. doi: 10.1016/S0140-6736(12)61728-0

Lyu, D., Msimbira, L. A., Nazari, M., Antar, M., Pagé, A., Shah, A., et al. (2021). The coevolution of plants and microbes underpins sustainable agriculture. *Microorganisms* 9:1036. doi: 10.3390/microorganisms9051036

Neal, A. L., Ahmad, S., Gordon-Weeks, R., and Ton, J. (2012). Benzoxazinoids in root exudates of maize attract pseudomonas putida to the rhizosphere. *PLoS One* 7:e35498. doi: 10.1371/JOURNAL.PONE.0035498

Pande, V., Joshi, T., Pandey, S. C., Sati, D., Mathpal, S., Pande, V., et al. (2021). Molecular docking and molecular dynamics simulation approaches for evaluation of laccase-mediated biodegradation of various industrial dyes. *J. Biomol. Struct. Dyn.*, 1–11. doi: 10.1080/07391102.2021.1971564

Papenfort, K., and Bassler, B. L. (2016). Quorum sensing signal–response systems in gram-negative bacteria. *Nat. Rev. Microbiol.* 14, 576–588. doi: 10.1038/NRMICRO.2016.89

Patel, H. K., Suárez-Moreno, Z. R., Degrassi, G., Subramoni, S., González, J. F., and Venturi, V. (2013). Bacterial LuxR solos have evolved to respond to different molecules including signals from plants. *Front. Plant Sci.* 4:447. doi: 10.3389/fpls.2013.00447

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., et al. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* 29, 845–854. doi: 10.1093/BIOINFORMATICS/BTT055

Sargsyan, K., Grauffel, C., and Lim, C. (2017). How molecular size impacts RMSD applications in molecular dynamics simulations. *J. Chem. Theory Comput.* 13, 1518–1524. doi: 10.1021/ACS.JCTC.7B00028/SUPPL_FILE/CT7B00028_SI_001.PDF

Sasse, J., Martinoia, E., and Northen, T. (2018). Feed your friends: do plant exudates shape the root microbiome? *Trends Plant Sci.* 23, 25–41. doi: 10.1016/j.tplants.2017.09.003

Sati, D., Pande, V., Pandey, S. C., and Samant, M. (2022). Recent advances in PGPR and molecular mechanisms involved in drought stress resistance. *J. Soil Sci. Plant Nutr.*, 1–19. doi: 10.1007/S42729-021-00724-5

Schandry, N., and Becker, C. (2020). Allelopathic plants: models for studying plant–interkingdom interactions. *Trends Plant Sci.* 25, 176–185. doi: 10.1016/j.tplants.2019.11.004

Subramoni, S., and Venturi, V. (2009). LuxR-family 'solos': bachelor sensors/regulators of signalling molecules. *Microbiology* 155, 1377–1385. doi: 10.1099/mic.0.026849-0

Subramoni, S., Gonzalez, J. F., Johnson, A., Péchy-Tarr, M., Rochat, L., Paulsen, I., et al. (2011). Bacterial subfamily of LuxR regulators that respond to plant compounds. *Appl. Environ. Microbiol.* 77, 4579–4588. doi: 10.1128/AEM.00183-11

Trott, O., and Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461. doi: 10.1002/JCC.21334

Wallner, B., and Elofsson, A. (2003). Can correct protein models be identified? *Protein Sci.* 12, 1073–1086. doi: 10.1110/ps.0236803

Wiederstein, M., and Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 35, W407–W410. doi: 10.1093/NAR/GKM290

Physiological, Structural, and Functional Insights Into the Cryoprotection of Membranes by the Dehydrins

Marijke R. Murray[1] and Steffen P. Graether[1,2]*

[1]Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada, [2]Graduate Program in Bioinformatics, University of Guelph, Guelph, ON, Canada

Plants can be exposed to cold temperatures and have therefore evolved several mechanisms to prevent damage caused by freezing. One of the most important targets are membranes, which are particularly susceptible to cold damage. To protect against such abiotic stresses, plants express a family of proteins known as late embryogenesis abundant (LEA) proteins. Many LEA proteins are intrinsically disordered, that is, they do not contain stable secondary or tertiary structures alone in solution. These proteins have been shown in a number of studies to protect plants from damage caused by cold, drought, salinity, and osmotic stress. In this family, the most studied proteins are the type II LEA proteins, better known as dehydrins (dehydration-induced proteins). Many physiological studies have shown that dehydrins are often located near the membrane during abiotic stress and that the expression of dehydrins helps to prevent the formation of oxidation-modified lipids and reduce the amount of electrolyte leakage, two hallmarks of damaged membranes. One of the earliest biophysical clues that dehydrins are involved in membrane cryoprotection came from *in vitro* studies that demonstrated a binding interaction between the protein and membranes. Subsequent work has shown that one conserved motif, known as K-segments, is involved in binding, while recent studies have used NMR to explore the residue specific structure of dehydrins when bound to membranes. The biophysical techniques also provide insight into the mechanism by which dehydrins protect the membrane from cold stress, which appears to mainly involve the lowering of the transition temperature.

Keywords: dehydrins, membranes, abiotic stress, physiological response, intrinsically disordered proteins, structure

INTRODUCTION

The plasma membrane is a primary site of damage resulting from stressors such as cold, drought, and osmotic stress. Each of these stressors cause a host of structural and compositional changes to the membrane lipids (Gupta et al., 2016). In this review, we mainly focus on the damage effects caused by low temperatures, and the evidence that dehydrins are involved in

preventing that damage. During periods of cold temperatures, the degree of unsaturation lipids in the cell membranes increases. Conversely, during warm temperatures, the membrane lipids become more saturated (Murata and Los, 1997). This changes the membrane fluidity, which induces a conformational change in many transmembrane and peripheral proteins. Membrane receptors then respond to the stress-induced change by initiating signaling pathways that influence gene expression related to stress tolerance (Kosová et al., 2015). This initial stress response is called the alarm phase and involves several proteomic and chemical changes to the cell environment, including phosphorylation of phosphoinositides (PIs) and enzymes in their metabolic pathways (Kosová et al., 2015).

Despite the desaturation of membrane lipids, there is still the danger of cold stress inhibiting normal membrane function (Steponkus, 1984). For example, dehydration caused by freezing temperatures can result in a phenomenon known as interbilayer fusion, which occurs between the plasma membrane and other intracellular membranes, such as mitochondrial, chloroplast, or Golgi membranes (Welti et al., 2002). At temperatures from −4°C to −10°C, membranes can undergo lamellar to hexagonal (II) phase transitions (Welti et al., 2002). This phase transition is a shift from a typical lipid bilayer to an inverted micelle structure with the fatty acid tails facing outward (Jouhet, 2013). Although the lamellar to hexagonal (II) phase is reversible, both of these occurrences can have serious effects on plant health and survival.

DEHYDRIN—A PLANT STRESS RESPONSE PROTEIN

One protective method that plants have evolved to combat this damage involves the expression of proteins known as dehydration-induced proteins (dehydrins). Dehydrins are a group of proteins expressed in plants to protect them from damage caused by dehydration stress (mainly drought, cold, salinity, and osmotic stresses; Graether and Boddington, 2014; Aziz et al., 2021; Riyazuddin et al., 2021; Smith and Graether, 2022). Dehydrins were originally identified as being group II proteins of the late embryogenesis abundant (LEA) proteins, a large family of proteins that are expressed in response to abiotic stress. LEA proteins were first recognized as being overexpressed in seeds (Galau et al., 1986; Roberts et al., 1993), and they have since been found to be expressed under abiotic stress conditions in adult plants as well. The presence of dehydrins has been established in the genomes of all embryophytes (Malik et al., 2017; Nagaraju et al., 2018), suggesting that they have been around at least since the emergence of early land plants, and may possibly predate them (Li et al., 1998).

Dehydrins are defined by the presence of one or more lysine-rich, semi-conserved motifs known as K-segments (Close, 1997). Previous work (Malik et al., 2017) has shown that the K-segment can be defined as [XKXGXX(D/E)KIK(D/E)KXPG], where X can be any amino acid. The conservation of the lysine residue positions suggests that this amino acid plays a

key role in dehydrin function, which was shown to be true for membrane protection (described below).

In addition to the K-segment, dehydrins also may contain three other motifs known as the Y-, F-, and S-segments. The motifs are present a variable number of times in dehydrins, resulting in several different architectures, specifically YSK_n, FK_n, FSK_n, SK_n, YK_n, K_n, and K_nS (Smith and Graether, 2022). In between, the conserved motifs are what are known as ϕ-segments; poorly conserved regions that mainly consist of small amino acids and are highly variable in length. Another conserved motif in dehydrin is the S-segment. As the name implies, this motif contains serine residues and can be described as [LHR(S/T)GS$_{4-6}$(S/D/E)(D/E)$_3$] (Malik et al., 2017). Two possible roles have been identified for this segment, both of which involve phosphorylation of the serines (Goday et al., 1994; Alsheikh et al., 2003). One study showed that the phosphorylation helps promote localization of the dehydrin to the nucleus (Goday et al., 1994), where it may be able to protect DNA and nuclear proteins. Another suggests that they are involved in the binding of calcium ions, which the authors speculate may allow the dehydrin to buffer the intracellular calcium concentration (Alsheikh et al., 2003).

The Y-segment, which can be described as [D(D/E)(Y/H/F) GNPX], where the X is often a hydrophobic amino acid, was originally named after the conserved tyrosine residue in the middle of the motif. Further analysis shows that this position may instead contain histidine or phenylalanine. Note that tryptophan, while aromatic is not found at this position, likely because compared to other aromatic amino acids, it promotes folding (Barua et al., 2008; Ghanmi et al., 2022). The F-segment ([EXXDRGXFDFX(G/K)]) is a relatively recently identified motif (Strimbeck, 2017; Riley et al., 2019). Similar to the Y-segment, it is named after the presence of two conserved aromatic amino acids, namely, phenylalanine. The biological roles of these two motifs have not yet been fully identified. Previously, it was speculated that the Y-segment may bind ATP, but this was shown to not be correct (Boddington and Graether, 2019).

Dehydrin sequences are rich in polar and charged amino acids, and therefore, it is not surprising that they are intrinsically disordered proteins (IDPs; Tompa, 2002; Uversky, 2002). IDPs are characterized as proteins that do not maintain stable 2D or 3D structures and instead are better described as having an extended coil structure. In the case of dehydrins, circular dichroism and NMR studies have shown that when the protein is alone in solution, it does not have an overall 3D shape, instead consisting mostly of coil structure, with a weak propensity for helical character in the K-segments (Findlater and Graether, 2009; Ágoston et al., 2011; Hughes and Graether, 2011).

LOCALIZATION OF DEHYDRINS TO MEMBRANES

Localization of proteins within the cell is often one of the first steps in determining the biological role of proteins. Many studies have shown that dehydrins localize to cellular

membranes, suggesting that they could protect them. One of the earliest studies followed dehydrin localization in *Zea mays* using immunogold-labeled antibodies that had been raised against the K-segment (Egerton-Warburton et al., 1997). The gold particles were found dispersed throughout the cytosol, nucleus, but were localized in an uneven manner near membranes. This unevenness was not observed near the endoplasmic reticulum, indicating that they may be preferentially binding some membranes but not others (Egerton-Warburton et al., 1997). Another study on the wheat dehydrin WCOR410 found that it localizes near the plasma membrane in cold-acclimated plants (Danyluk et al., 1998). Under closer inspection, they found that immunogold particles were highly concentrated around the fibrillar network between cell walls, despite not having a known signal peptide for extracellular transport (Danyluk et al., 1998). Taken together, these two studies reveal the potential importance dehydrins in protecting membranes.

More recent studies using subcellular fractionation also provide insight into dehydrins function. The *Solanum sogarandinum* dehydrin DHN24 was also found to be membrane-localized in various tissue types (Szabala et al., 2014). At a subcellular level, immunogold-labeled DHN24 localized to the outer mitochondrial membrane of cold-acclimated phloem cells. The dehydrin was not detected in the mitochondrial component after performing subcellular fractionation, which indicates that DHN24 may bind the outer mitochondrial membrane from the cytosolic side but is not transported into the organelle (Szabala et al., 2014). Cellular fractionation experiments on celery petioles found the calcium-binding, dehydrin-like protein VCaB45 associated with vacuole membranes (Heyen et al., 2002). They treated vacuole membranes with Triton X-100 detergent to show that the protein was bound to the luminal side of the membrane, requiring some sort of import mechanism (Heyen et al., 2002).

Studies on dehydrin localization in transgenic plants enable the researcher to have a greater level of control over experimental conditions and dehydrin type. It has been found that cold acclimation played a role in dehydrin localization to the cell membranes (Puhakainen et al., 2004). Two-thirds of the immunogold-labeled LTI29 were found in the cytosol before acclimation. After cold acclimation, 2/3 of the gold labeling was found on cell membranes and only 1/3 throughout the cytosol, strongly implicating dehydrins in cryoprotection of the plasma membrane (Puhakainen et al., 2004). Another study identified potential protein–protein interactions at the membrane of the pepper dehydrin CaDHN3 (Meng et al., 2021). CaDHN3-GFP fusion protein was localized to the nucleus and cell membrane of *N. benthamiana* leaves. Yeast two hybrid experiments showed that CaDHN3 was found to interact with CaHIRD11, another dehydrin, at the plasma membrane (Meng et al., 2021). A recent bimolecular fluorescent complementation assay demonstrated an association between a dehydrin and aquaporin, a key membrane protein for controlling intracellular water content (Hernández-Sánchez et al., 2019). This opens the possibility of multiple potential dehydrin-dehydrin and dehydrin-membrane protein interactions and may have a role

in the mechanism behind membrane stress protection (Hernández-Sánchez et al., 2019; Meng et al., 2021).

PHYSIOLOGICAL PROTECTION BY DEHYDRINS

One major complication of freezing stress in plants is the buildup of reactive oxygen species causing lipid peroxidation, which decreases membrane fluidity (Barclay and McKersie, 1994). Two effective methods to gauge membrane damage are measuring malondialdehyde (MDA) levels (MDA is the final product of lipid peroxidation), and measuring electrolyte leakage (EL), since the plasma membrane loses structure under stress, allowing ions to cross the membrane in higher quantities. Conifers represent an interesting model for studying physiological protection from dehydrins because of their natural ability to tolerate extreme cold. It has been observed that the relative electrolyte leakage in cold-acclimated boreal conifers was so low that cells only sustained reversible damage (Strimbeck et al., 2008). To expand upon these results, another study measured dehydrin expression throughout the year in Siberian spruce; Western blot analysis revealed peak dehydrin expression during the winter months when cold protection is needed the most (Kjellsen et al., 2013). The band intensity of dehydrins coincided with T_m values, the midpoint monthly temperature; when T_m decreased, band intensity increased. In other words, dehydrin expression was highest when the temperature was lowest (Kjellsen et al., 2013). One study measured MDA levels in *Eriobotrya japonica* subjected to freezing stress in the presence of a range of different dehydrins (Xu et al., 2014). Following the stress treatment, MDA levels were 56% higher in freeze intolerant plants than freeze tolerant plants, which is correlated with lower levels of dehydrins in the susceptible cultivar (Xu et al., 2014). This study showed that increased dehydrin expression is implicated in freeze tolerance and decreased levels of oxidative stress.

Dehydrin overproduction in transgenic crops is not just an excellent way to corroborate results seen in native, cold-tolerant plants but is also a promising avenue to extend the growing season and protect plants from extreme weather events. Transgenic strawberry lines overexpressing the wheat dehydrin gene *Wcor410a* were used to test the cryoprotective properties of dehydrins (Houde et al., 2004). The cold-acclimated, transgenic strawberry plants showed 20% EL compared to 60% in the acclimated wild type, and also demonstrated better tolerance to sub-zero temperatures. Consistent with earlier studies performed on transgenic plants, tobacco overexpressing the citrus dehydrin CuCOR19 had decreased levels of EL, indicating less membrane damage than the wild type (Hara et al., 2003). Interestingly, the transgenic tobacco also exhibited faster growth at milder temperatures than the plants not expressing CuCOR19, suggesting that the protective role of dehydrins may not only occur at freezing temperatures. This idea is supported by a study on cultivated tomato, which is known to have few genes related to environmental stress response. *Solanum habrochaites* is a variety of wild tomato that exhibits increased tolerance

to the cold (Liu et al., 2015). In the study, they created a transgenic tomato plant by expressing the *S. habrochaites* dehydrin, ShDhn, in the cultivated tomato. Compared to the control wild type, the transgenic lines retained more turgidity after being subjected to a 4°C treatment for 3 days. Following the cold stress, EL of the wild type increased 3.3-fold, while the transgenic lines increased only 2.6-fold. After a 10-day period of drought stress, the transgenic plants retained stem turgidity, while the wild type lost the ability to stand upright (Liu et al., 2015).

STRUCTURAL CHANGES IN DEHYDRIN UPON MEMBRANE BINDING

IDPs are not necessarily fully disordered all of the time, many gain structure in the presence of a ligand. This is also the case for dehydrins, and some of the earliest biophysical evidence for a role for dehydrins in protecting membranes came from structural experiments. CD experiments showed that the 35 kDa cowpea dehydrin gained α-helical structure in the presence of SDS micelles (Ismail et al., 1999). These micelles are often used as membrane mimetics because they are easier to use with biophysical techniques while still being a good model system for biological lipids in membranes (Tulumello and Deber, 2009). Subsequent studies have shown that the K-segment is involved in SDS binding and that the gain in α-helicity occurs in these regions (Koag et al., 2009; Atkinson et al., 2016). The deletion of one or two of the K-segments of *Zea mays* DHN1 resulted in less gain in α-helicity and that the deletion of all three K-segments resulted in no structural change (Koag et al., 2009). A later study used NMR to probe the structure of the K_2 dehydrin in the presence of SDS micelles, obtaining structural information on a per residue basis (Atkinson et al., 2016). This study showed that the K_2 dehydrin, a protein found in *Vitis riparia* that is a splice variant of the longer YSK_2 dehydrin, also gained α-helical structure when bound to SDS. Dynamics data from the K_2 study showed that the bound K-segments were less disordered, while the ϕ-segments (i.e., the poorly conserved region of the dehydrins that connects the conserved segments) retained their flexibility. This supports the idea that the ϕ-segments are highly flexible, even within the context of a disordered protein, and allow for the conserved segments to optimally orient themselves to interact with the ligand(s; Hughes and Graether, 2011).

A more detailed structural analysis (Clarke et al., 2015) using the second structure propensity program δ2Δ (Camilloni et al., 2012) and NMR dynamics experiments showed that the middle of the K-segment was helical nearly 80–90% of the time, while the 3–4 residues flanking these were helical ~20–60% of the time. Extensive modeling of the K_2-SDS micelle interaction suggested that the Lys residues flank the hydrophobic and negatively charged sides of the K-segment helices, while the few hydrophobic residues of the K-segment are buried near the acyl chains of the membrane (Clarke et al., 2015). A similar result was shown from NMR data using *Arabidopsis thaliana* Lti30 K-segments while bound to small unilamellar vesicle
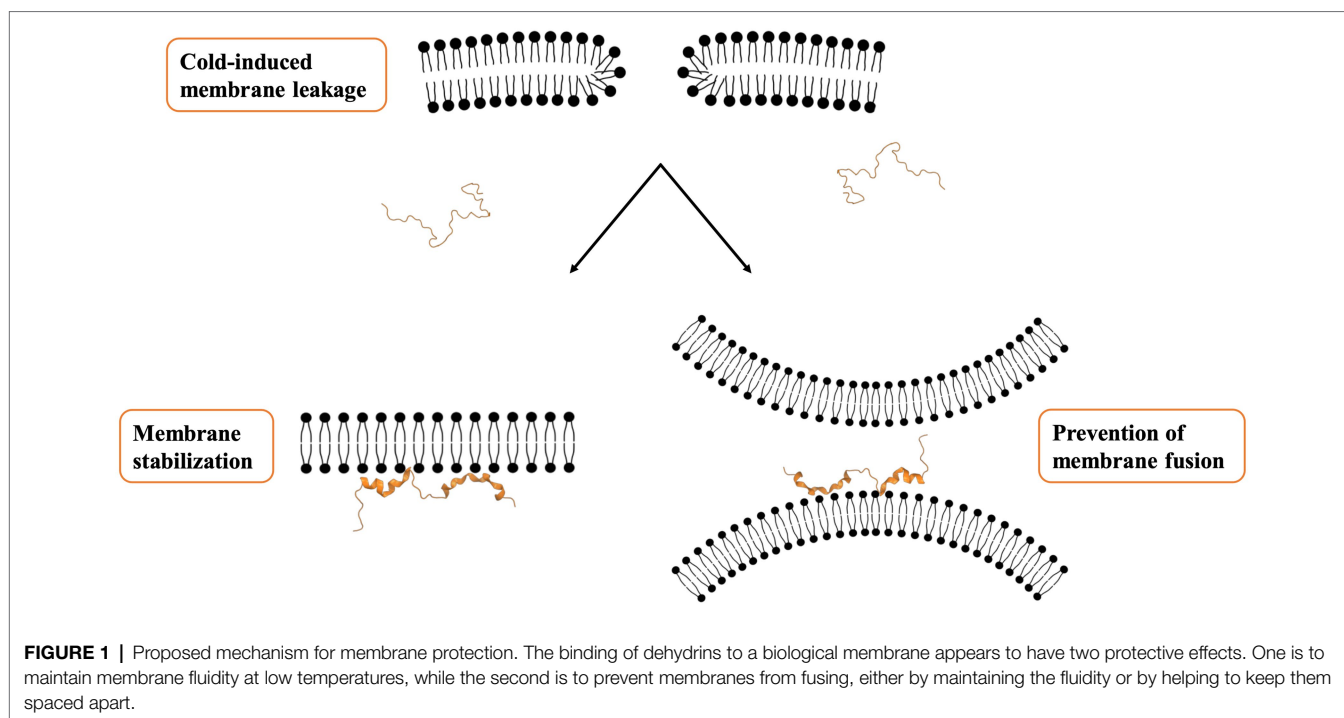
(SUVs, also referred to as liposomes; Eriksson et al., 2016). The center of the peptides was highly helical, though the flanking residues were in an extended, β-strand conformation. The author modeled the interaction between the protein and membrane, and also suggested that the polar and negatively charged face of the K-segment helix are oriented away from the membrane surface (Eriksson et al., 2016).

Other studies have also used SUVs to see whether dehydrins are able to interact with membranes that more closely resemble a biological one. Most studies found that dehydrin largely bound to anionic lipids in SUVs and that an increase in α-helicity was also observed, though at a lower level compared to the proteins in the presence of micelles (Soulages et al., 2003; Koag et al., 2009; Clarke et al., 2015). The requirement for a negative charge on the lipid headgroups likely reflects binding by the Lys residues in the K-segments (Clarke et al., 2015; Atkinson et al., 2016). The ionization state of the di-His sequences that flank some K-segments are also important for binding (Eriksson et al., 2011). Using the Lti30 dehydrin, they showed that the deprotonation of the His residues prevented binding of the protein to SUVs. These His residues are not part of the K-segment helix but may be positioning themselves so that they interact with the amines on the head groups (Eriksson et al., 2011).

The need for anionic headgroups appears to depend on the dehydrin and the lipids being used in the study. In several of the studies reported above, it was shown that SUVs consisting only of phosphatidylcholine (PC) did not bind dehydrins (Soulages et al., 2003; Koag et al., 2009; Clarke et al., 2015). PC is zwitterionic, but the presence of the negative phosphate in the headgroup did not appear to be sufficient to allow for binding as measured by CD experiments or SUV pulldown assays. However, other dehydrin studies appear to contradict either the change in structure or the need for acidic lipids. Experiments using ERD10 and ERD14, dehydrins found in *Arabidopsis thaliana*, did not show a change in the CD spectra in the presence of acidic SUVs (i.e., there was no evidence of a gain in α-helicity), yet mini-gel filtration assays suggest that the proteins did interact (Kovacs et al., 2008). Likewise, using surface plasmon resonance, an interaction was also detected between Lti30 and PC SUVs (Eriksson et al., 2011). Whether these results come from a specific feature of these dehydrins, or whether detection depends on the exact experimental conditions and measuring equipment used, will require further studies.

MECHANISTIC EVIDENCE FOR MEMBRANE PROTECTION

In addition to monitoring structural changes in dehydrins, biophysical experiments also allow one to explore the mechanisms by which dehydrins exert their protective role on membranes (**Figure 1**). Studies have brought protective mechanisms to light, with two effects being reported; one is the decrease in the membrane transition temperature (T_m; Clarke et al., 2015), and the other a lowering of the relative

FIGURE 1 | Proposed mechanism for membrane protection. The binding of dehydrins to a biological membrane appears to have two protective effects. One is to maintain membrane fluidity at low temperatures, while the second is to prevent membranes from fusing, either by maintaining the fluidity or by helping to keep them spaced apart.

humidity (R_h) at which the transition to hexagonal (II) phase occurs, where the membrane forms an inverted micelle structure (Andersson et al., 2020). Both share the common effect of maintaining membrane fluidity and keep it functional despite changes in temperature or water content. In the study with the *Vitis riparia* K_2, it was also found that the protein was able to prevent the fusion of liposomes after a freeze/thaw stress was applied (Clarke et al., 2015). Specifically, there was a correlation between the reduction in the average size of the fused SUVs (as assessed by dynamic light scattering) and the concentration of K_2. The use of PEG3350 (a polymer with a similar hydrodynamic radius to K_2) did not prevent fusion, while the addition of K-segment peptides alone (i.e., a dehydrin without its φ-segments) did prevent it. These two control experiments suggest that the dehydrins prevent fusion not by steric hindrance but by altering the fluidity of the membrane through a direct interaction. A recent study showed that K-segments alone were also able to prevent aggregation of liposomes and that removing any of the charged or hydrophobic residues in the sequence drastically inhibited this protective function (Kimura et al., 2022). While membrane fusion was not directly examined, a different dehydrin (Lti30) using another assay showed that these proteins were able to cluster SUVs. Their interpretation of the role of the dehydrin is to maintain a consistent separation of membranes, where one K-segment in Lti30 would bind one membrane, while another K-segment could bind another membrane (Eriksson et al., 2016). As water content might vary, for example, during cold stress, these dehydrins would provide the structure needed to keep membrane integrity and prevent them from rupturing. In this model, the φ-segments could be extended so as to facilitate the "cross-linking" of two different SUVs.

CONCLUSION

Dehydrin-membrane binding plays a fundamental role in plant cryoprotection, as membranes are one of the primary sites of cold-induced damage. One issues that need resolving is the multiple protective functions of dehydrins in addition to membrane protection (e.g., enzyme protection and protecting DNA from reactive oxygen species). While moonlighting has shown that IDPs can have multiple functions, why a dehydrin can switch between membrane protection at the membrane, enzyme protection in the cytosol, and DNA protection in the nucleus is not yet clear. In addition, the exact mechanism by which cryoprotection occurs will need further study, because different effects have been described by various studies. Dehydrins can reduce the amount of membrane fusion, while they can also promote membrane clustering. Two different protective mechanisms have also been proposed as: keeping membranes fluid versus keeping them consistently separated. To answer, all of these questions may require studies with *in vivo* systems, where there will be direct evidence of dehydrin cryoprotection and membrane binding in the cell.

AUTHOR CONTRIBUTIONS

MM and SG wrote and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

REFERENCES

Ágoston, B. S., Kovacs, D., Tompa, P., and Perczel, A. (2011). Full backbone assignment and dynamics of the intrinsically disordered dehydrin ERD14. *Biomol. NMR Assign.* 5, 189–193. doi: 10.1007/s12104-011-9297-2

Alsheikh, M. K., Heyen, B. J., and Randall, S. K. (2003). Ion binding properties of the dehydrin ERD14 are dependent upon phosphorylation. *J. Biol. Chem.* 278, 40882–40889. doi: 10.1074/jbc.M307151200

Andersson, J. M., Pham, Q. D., Mateos, H., Eriksson, S., Harryson, P., and Sparr, E. (2020). The plant dehydrin Lti30 stabilizes lipid lamellar structures in varying hydration conditions. *J. Lipid Res.* 61, 1014–1024. doi: 10.1194/jlr.ra120000624

Atkinson, J., Clarke, M. W., Warnica, J. M., Boddington, K. F., and Graether, S. P. (2016). Structure of an intrinsically disordered stress protein alone and bound to a membrane surface. *Biophys. J.* 111, 480–491. doi: 10.1016/j.bpj.2016.07.001

Aziz, M. A., Sabeem, M., Mullath, S. K., Brini, F., and Masmoudi, K. (2021). Plant group II LEA proteins: intrinsically disordered structure for multiple functions in response to environmental stresses. *Biomol. Ther.* 11:1662. doi: 10.3390/biom11111662

Barclay, K. D., and McKersie, B. D. (1994). Peroxidation reactions in plant membranes: effects of free fatty acids. *Lipids* 29, 877–882. doi: 10.1007/BF02536256

Barua, B., Lin, J. C., Williams, V. D., Kummler, P., Neidigh, J. W., and Andersen, N. H. (2008). The Trp-cage: optimizing the stability of a globular miniprotein. *Protein Eng. Des. Sel.* 21, 171–185. doi: 10.1093/protein/gzm082

Boddington, K. F., and Graether, S. P. (2019). Binding of a Vitis riparia Dehydrin to DNA. *Plant Sci. Int. J. Exp. Plant Biol.* 287:110172. doi: 10.1016/j.plantsci.2019.110172

Camilloni, C., Simone, A. D., Vranken, W. F., and Vendruscolo, M. (2012). Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry* 51, 2224–2231. doi: 10.1021/bi3001825

Clarke, M. W., Boddington, K. F., Warnica, J. M., Atkinson, J., McKenna, S., Madge, J., et al. (2015). Structural and functional insights into the Cryoprotection of membranes by the intrinsically disordered Dehydrins. *J. Biol. Chem.* 290:678219. doi: 10.1074/jbc.m115.678219

Close, T. J. (1997). Dehydrins: A commonalty in the response of plants to dehydration and low temperature. *Physiol. Plant.* 100, 291–296. doi: 10.1111/j.1399-3054.1997.tb04785.x

Danyluk, J., Perron, A., Houde, M., Limin, A., Fowler, B., Benhamou, N., et al. (1998). Accumulation of an acidic dehydrin in the vicinity of the plasma membrane during cold acclimation of wheat. *Plant Cell* 10, 623–638. doi: 10.1105/tpc.10.4.623

Egerton-Warburton, L. M., Balsamo, R. A., and Close, T. J. (1997). Temporal accumulation and ultrastructural localization of dehydrins in Zea mays. *Physiol. Plant.* 101, 545–555. doi: 10.1111/j.1399-3054.1997.tb01036.x

Eriksson, S., Eremina, N., Barth, A., Danielsson, J., and Harryson, P. (2016). Membrane-induced folding of the plant-stress protein Lti30. *Plant Physiol.* 171, 932–943. doi: 10.1104/pp.15.01531

Eriksson, S. K., Kutzer, M., Procek, J., Gröbner, G., and Harryson, P. (2011). Tunable membrane binding of the intrinsically disordered dehydrin Lti30, a cold-induced plant stress protein. *Plant Cell* 23, 2391–2404. doi: 10.1105/tpc.111.085183

Findlater, E. E., and Graether, S. P. (2009). NMR assignments of the intrinsically disordered K2 and YSK2 dehydrins. *Biomol. NMR Assign.* 3, 273–275. doi: 10.1007/s12104-009-9192-2

Galau, G. A., Hughes, D. W., and Dure, L. (1986). Abscisic acid induction of cloned cotton late embryogenesis-abundant (lea) mRNAs. *Plant Mol. Biol.* 7, 155–170. doi: 10.1007/BF00021327

Ghanmi, S., Graether, S. P., and Hanin, M. (2022). The halophyte Dehydrin sequence landscape. *Biomol. Ther.* 12:330. doi: 10.3390/biom12020330

Goday, A., Jensen, A. B., Culiáñez-Macià, F. A., Albà, M. M., Figueras, M., Serratosa, J., et al. (1994). The maize abscisic acid-responsive protein Rab17 is located in the nucleus and interacts with nuclear localization signals. *Plant Cell* 6, 351–360. doi: 10.1105/tpc.6.3.351

Graether, S. P., and Boddington, K. F. (2014). Disorder and function: a review of the dehydrin protein family. *Front. Plant Sci.* 5:576. doi: 10.3389/fpls.2014.00576

Gupta, D. B., Rai, Y., Gayali, S., Chakraborty, S., and Chakraborty, N. (2016). Plant Organellar proteomics in response to dehydration: turning protein repertoire into insights. *Front. Plant Sci.* 7:460. doi: 10.3389/fpls.2016.00460

Hara, M., Terashima, S., Fukaya, T., and Kuboi, T. (2003). Enhancement of cold tolerance and inhibition of lipid peroxidation by citrus dehydrin in transgenic tobacco. *Planta* 217, 290–298. doi: 10.1007/s00425-003-0986-7

Hernández-Sánchez, I. E., Maruri-López, I., Molphe-Balch, E. P., Becerra-Flora, A., Jaimes-Miranda, F., and Jiménez-Bremont, J. F. (2019). Evidence for in vivo interactions between dehydrins and the aquaporin AtPIP2B. *Biochem Bioph Res Co* 510, 545–550. doi: 10.1016/j.bbrc.2019.01.095

Heyen, B. J., Alsheikh, M. K., Smith, E. A., Torvik, C. F., Seals, D. F., and Randall, S. K. (2002). The calcium-binding activity of a vacuole-associated, dehydrin-like protein is regulated by phosphorylation. *Plant Physiol.* 130, 675–687. doi: 10.1104/pp.002550

Houde, M., Dallaire, S., N'Dong, D., and Sarhan, F. (2004). Overexpression of the acidic dehydrin WCOR410 improves freezing tolerance in transgenic strawberry leaves. *Plant Biotechnol. J.* 2, 381–387. doi: 10.1111/j.1467-7652.2004.00082.x

Hughes, S., and Graether, S. P. (2011). Cryoprotective mechanism of a small intrinsically disordered dehydrin protein. *Protein Sci. Pub. Protein Soc.* 20, 42–50. doi: 10.1002/pro.534

Ismail, A. M., Hall, A. E., and Close, T. J. (1999). Purification and partial characterization of a Dehydrin involved in chilling tolerance during seedling emergence of cowpea. *Plant Physiol.* 120, 237–244. doi: 10.1104/pp.120.1.237

Jouhet, J. (2013). Importance of the hexagonal lipid phase in biological membrane organization. *Front. Plant Sci.* 4:494. doi: 10.3389/fpls.2013.00494

Kimura, Y., Ohkubo, T., Shimizu, K., Magata, Y., Park, E. Y., and Hara, M. (2022). Inhibition of cryoaggregation of phospholipid liposomes by an Arabidopsis intrinsically disordered dehydrin and its K-segment. *Coll. Surf. B. Biointer.* 211:112286. doi: 10.1016/j.colsurfb.2021.112286

Kjellsen, T. D., Yakovlev, I. A., Fossdal, C. G., and Strimbeck, G. R. (2013). Dehydrin accumulation and extreme low-temperature tolerance in Siberian spruce (*Picea obovata*). *Tree Physiol.* 33, 1354–1366. doi: 10.1093/treephys/tpt105

Koag, M.-C., Wilkens, S., Fenton, R. D., Resnik, J., Vo, E., and Close, T. J. (2009). The K-segment of maize DHN1 mediates binding to anionic phospholipid vesicles and concomitant structural changes. *Plant Physiol.* 150, 1503–1514. doi: 10.1104/pp.109.136697

Kosová, K., Vítámvás, P., Urban, M. O., Klíma, M., Roy, A., and Prášil, I. T. (2015). Biological networks underlying abiotic stress tolerance in temperate crops—A proteomic perspective. *Int. J. Mol. Sci.* 16, 20913–20942. doi: 10.3390/ijms160920913

Kovacs, D., Kalmar, E., Torok, Z., and Tompa, P. (2008). Chaperone activity of ERD10 and ERD14, two disordered stress-related plant proteins. *Plant Physiol.* 147, 381–390. doi: 10.1104/pp.108.118208

Li, R., Brawley, S. H., and Close, T. J. (1998). Proteins immunologically related to dehydrins in fucoid algae. *J. Phycol.* 34, 642–650. doi: 10.1046/j.1529-8817.1998.340642.x

Liu, H., Yu, C., Li, H., Ouyang, B., Wang, T., Zhang, J., et al. (2015). Overexpression of ShDHN, a dehydrin gene from Solanum habrochaites enhances tolerance to multiple abiotic stresses in tomato. *Plant Sci.* 231, 198–211. doi: 10.1016/j.plantsci.2014.12.006

Malik, A. A., Veltri, M., Boddington, K. F., Singh, K. K., and Graether, S. P. (2017). Genome analysis of conserved Dehydrin motifs in vascular plants. *Front. Plant Sci.* 8:709. doi: 10.3389/fpls.2017.00709

Meng, Y.-C., Zhang, H.-F., Pan, X.-X., Chen, N., Hu, H.-F., Haq, S., et al. (2021). CaDHN3, a pepper (*Capsicum annuum* L.) Dehydrin gene enhances the tolerance against salt and drought stresses by reducing ROS accumulation. *Int. J. Mol. Sci.* 22:3205. doi: 10.3390/ijms22063205

Murata, N., and Los, D. A. (1997). Membrane fluidity and temperature perception. *Plant Physiol.* 115, 875–879. doi: 10.1104/pp.115.3.875

Nagaraju, M., Reddy, P. S., Kumar, S. A., Kumar, A., Suravajhala, P., Ali, A., et al. (2018). Genome-wide in silico analysis of dehydrins in Sorghum bicolor, Setaria italica and Zea mays and quantitative analysis of dehydrin gene expressions under abiotic stresses in Sorghum bicolor. *Plant Gene* 13, 64–75. doi: 10.1016/j.plgene.2018.01.004

Puhakainen, T., Hess, M. W., Mäkelä, P., Svensson, J., Heino, P., and Palva, E. T. (2004). Overexpression of multiple dehydrin genes enhances tolerance to

freezing stress in Arabidopsis. *Plant Mol. Biol.* 54, 743–753. doi: 10.1023/B:PLAN.0000040903.66496.a4

Riley, A. C., Ashlock, D. A., and Graether, S. P. (2019). Evolution of the modular, disordered stress proteins known as dehydrins. *PLoS One* 14:e0211813. doi: 10.1371/journal.pone.0211813

Riyazuddin, R., Nisha, N., Singh, K., Verma, R., and Gupta, R. (2021). Involvement of dehydrin proteins in mitigating the negative effects of drought stress in plants. *Plant Cell Rep.* 41, 519–533. doi: 10.1007/s00299-021-02720-6

Roberts, J. K., DeSimone, N. A., Lingle, W. L., and Dure, L. (1993). Cellular concentrations and uniformity of cell-type accumulation of two lea proteins in cotton embryos. *Plant Cell* 5, 769–780. doi: 10.2307/3869614

Smith, M. A., and Graether, S. P. (2022). The disordered Dehydrin and its role in plant protection: A biochemical perspective. *Biomol. Ther.* 12:294. doi: 10.3390/biom12020294

Soulages, J. L., Kim, K., Arrese, E. L., Walters, C., and Cushman, J. C. (2003). Conformation of a group 2 late embryogenesis abundant protein from soybean. Evidence of poly (L-proline)-type II structure. *Plant Physiol.* 131, 963–975. doi: 10.1104/pp.015891

Steponkus, P. L. (1984). Role of the plasma membrane in freezing injury and cold acclimation. *Annu. Rev. Plant Physiol.* 35, 543–584. doi: 10.1146/annurev.pp.35.060184.002551

Strimbeck, G. R. (2017). Hiding in plain sight: the F segment and other conserved features of seed plant SK$_n$ dehydrins. *Planta* 245, 1061–1066. doi: 10.1007/s00425-017-2679-7

Strimbeck, G. R., Kjellsen, T. D., Schaberg, P. G., and Murakami, P. F. (2008). Dynamics of low-temperature acclimation in temperate and boreal conifer foliage in a mild winter climate. *Tree Physiol.* 28, 1365–1374. doi: 10.1093/treephys/28.9.1365

Szabala, B. M., Fudali, S., and Rorat, T. (2014). Accumulation of acidic SK3 dehydrins in phloem cells of cold- and drought-stressed plants of the Solanaceae. *Planta* 239, 847–863. doi: 10.1007/s00425-013-2018-6

Tompa, P. (2002). Intrinsically unstructured proteins. *Trends Biochem. Sci.* 27, 527–533. doi: 10.1016/S0968-0004(02)02169-2

Tulumello, D. V., and Deber, C. M. (2009). SDS micelles as a membrane-mimetic environment for transmembrane segments. *Biochemistry* 48, 12096–12103. doi: 10.1021/bi9013819

Uversky, V. N. (2002). Natively unfolded proteins: A point where biology waits for physics. *Protein Sci.* 11, 739–756. doi: 10.1110/ps.4210102

Welti, R., Li, W. Q., Li, M. Y., Sang, Y. M., Biesiada, H., Zhou, H. E., et al. (2002). Profiling membrane lipids in plant stress responses - role of phospholipase D alpha in freezing-induced lipid changes in Arabidopsis. *J. Biol. Chem.* 277, 31994–32002. doi: 10.1074/jbc.M205375200

Xu, H., Yang, Y., Xie, L., Li, X., Feng, C., Chen, J., et al. (2014). Involvement of multiple types of Dehydrins in the freezing response in loquat (*Eriobotrya japonica*). *PLoS One* 9:e87575. doi: 10.1371/journal.pone.0087575

Structure-Based Modulation of the Ligand Sensitivity of a Tomato Dimeric Abscisic Acid Receptor Through a Glu to Asp Mutation in the Latch Loop

Lourdes Infantes[1†], Maria Rivera-Moreno[1†], Miguel Daniel-Mozo[1], Juan Luis Benavente[1], Javier Ocaña-Cuesta[2], Alberto Coego[2], Jorge Lozano-Juste[2], Pedro L. Rodriguez[2] and Armando Albert[1*]

[1] Instituto de Química Física Rocasolano, Consejo Superior de Investigaciones Científicas, Madrid, Spain, [2] Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas-Universidad Politécnica de Valencia, Valencia, Spain

The binding of the plant phytohormone Abscisic acid (ABA) to the family of ABA receptors (PYR/PYL/RCAR) triggers plant responses to abiotic stress. Thus, the implementation of genetic or chemical strategies to modulate PYR/PYL activity might be biotechnologically relevant. We have employed the available structural information on the PYR/PYL receptors to design SlPYL1, a tomato receptor, harboring a single point mutation that displays enhanced ABA dependent and independent activity. Interestingly, crystallographic studies show that this mutation is not directly involved in ABA recognition or in the downstream phosphatase (PP2C) inhibitory interaction, rather, molecular dynamic based ensemble refinement restrained by crystallographic data indicates that it enhances the conformational variability required for receptor activation and it is involved in the stabilization of an active form of the receptor. Moreover, structural studies on this receptor have led to the identification of niacin as an ABA antagonist molecule *in vivo*. We have found that niacin blocks the ABA binding site by mimicking ABA receptor interactions, and the niacin interaction inhibits the biochemical activity of the receptor.

Keywords: abiotic stress, abscisic acid, plant biology, protein crystallography, signal transduction, structural biology

INTRODUCTION

Abiotic stresses derived from global warming are threatening natural resources and food production; of them, drought is central as irrigated lands produce one-third of the world's food[1]. Hence, to improve crop production, it is sensible to understand the basis for plant adaptation to drought and the implementation of strategies to apply this knowledge. The effect of drought on plants is complex as they implement many protective adaptations, among them, stomata closure

---

[1] https://www.fao.org/sustainability/news/detail/en/c/1274219/

is key to prevent water evaporation and to maintain hydration of cells and tissues. This in turn, is controlled by the cytosolic levels of abscisic acid (ABA), the main phytohormone involved in the adaptive responses to drought (Cutler et al., 2010; Gonzalez-Guzman et al., 2012).

Abscisic acid signaling relies on the coordinated action of the family of pyrabactin resistance 1/PYR1-like/regulatory components of ABA receptors (PYR/PYL), on the clade A subfamily of protein phosphatases type-2C (PP2C) ABA coreceptors and on the family of sucrose non-fermenting-2-related protein kinases (SnRK2; Reviewed in Rodriguez et al., 2019). The available molecular and structural information on the components of the pathway provides a detailed mechanism for ABA sensing. In the resting state, the PYR/PYL receptors are monomers or dimers and display a wide empty ABA binding cavity flanked by two highly conserved loops, named as gate loop and latch loop. In this situation, the SnRK2s are forming a stable complex with the PP2Cs that maintains the kinase inactive. The increase in the ABA concentration motivated by stress gives rise to the hormone association with PYR/PYL receptors. This includes the structural rearrangement of the gate and latch loops in the ABA binding pocket and the dissociation of dimeric receptors to provide a binding surface for the PP2Cs. The formation of PYL-ABA-PP2Cs ternary complexes releases the SnRK2s from association with the PP2Cs (Melcher et al., 2009; Miyazono et al., 2009; Nishimura et al., 2009; Santiago et al., 2009; Yin et al., 2009; Dupeux et al., 2011a,b; Yunta et al., 2011). Then, the active SnRK2s can phosphorylate downstream effectors activating transcriptional and post-transcriptional ABA mediated cell response including stomata closure. ABA-activated SnRK2s that have been previously dephosphorylated by clade A PP2Cs precise activation by phosphorylation through B2 and B3-type RAF-like MAP3Ks (Lin et al., 2020, 2021; Takahashi et al., 2020).

The knowledge derived from the characterization of the ABA pathway has provided the basis for the genetic manipulation of plants for biotechnological purposes. For instance, the overexpression of ABA receptors (Santiago et al., 2009; Mosquna et al., 2011; Pizzio et al., 2013; Gonzalez-Guzman et al., 2014; Yang et al., 2016, 2019) or the reduced expression of the PP2C coreceptors (Rubio et al., 2009; Antoni et al., 2012) diminish plant transpiration and increase drought resistance. In addition, there is a field for the structure-based rational design of novel molecules that could be used as agrochemical compounds against stress. In particular, for the development of chemical compounds that act as ABA agonists or antagonists. These molecules can modulate ABA signaling dynamically and exogenously either by enhancing drought tolerance or inhibiting ABA mediated responses in plants, respectively, (Hewage et al., 2020; Lozano-Juste et al., 2020). Major breakthroughs in this field have been the discovery of Opabactin, a potent ABA agonist for manipulating crop abiotic stress tolerance and water use (Vaidya et al., 2019) and Antabactin, a pan-receptor antagonist that disrupts ABA mediated response in *Arabidopsis* and different crop species (Vaidya et al., 2021).

The advantages and challenges of genetic and chemical approaches have been reviewed (Yang et al., 2021), however, to pursue both approaches it would be desirable to study in depth the transitions leading to the activation of the pathway. The structure-based Gate-Latch-Lock mechanism for ABA signaling proposed that the latch and gate loops undergo a well-defined conformational change from the open conformation observed in the apo form of the PYR/PYL receptors to the closed conformation adopted in the ternary PYR/PYL-ABA-PP2C complexes (Melcher et al., 2009). This description of the steps leading to ABA perception was enriched by the identification and analysis of different intermediates (Moreno-Alvero et al., 2017). Thus, Moreno-Alvero et al. (2017) identified in crop ABA receptors a novel latch-closed gate-open ABA-bound intermediate, which suggests that PYR/PYL receptors display an equilibrium state between gate-open and gate-closed in the presence of ABA. These studies also revealed that latch closure precedes ABA binding, whereas gate closure follows the ABA binding event upon conformational selection by the PP2C. This highlights the role of PP2C as ABA coreceptors as PYR/PYL receptors display an equilibrium state between gate-open and gate-closed in the ABA-bound form and consequently, the PP2C either induces the shift of this equilibrium to the gate-closed form or will selectively bind to the gate-closed conformation. An additional step should be considered for the activation of dimeric PYR/PYLs as structural studies showed that the conformational changes leading to ABA recognition for these receptors are necessarily coupled with their dissociation because the dimerization and PP2C binding interfaces overlap. The energetic cost of such dissociation gives rise to lower intrinsic affinity for ABA, which is around two orders of magnitude lower than that of the monomeric receptors (Dupeux et al., 2011b). Moreover, monomeric receptors are able to inhibit PP2C activity to some extent in absence of ABA (Hao et al., 2011).

The relevance for ABA signaling of dimeric PYLs in tomato is presumably similar to that of dimeric PYLs in *Arabidopsis* (Sun et al., 2011; Cao et al., 2013; Okamoto et al., 2013; Gonzalez-Guzman et al., 2014; Chen et al., 2016). Phylogenetic studies have shown that tomato PYLs, as those from the higher plants (Yang et al., 2020), are grouped in three subfamilies, as *Arabidopsis* PYLs, and quinabactin, which is an agonist of dimeric ABA receptors, is able to inhibit tomato seed germination and induce stress-responsive genes (Gonzalez-Guzman et al., 2014). However, genetic studies where expression of tomato PYLs is modified are still scarce (Kai et al., 2019). *In vitro* studies have demonstrated that dimeric tomato PYLs are able to inhibit the activity of clade A PP2Cs, which play a key role for ABA signaling in tomato (Gonzalez-Guzman et al., 2014; Zhang et al., 2018).

Abscisic acid receptor agonists and antagonists are important tools to achieve either hyper- or hypostimulation of ABA signaling, respectively. Chemical activation of ABA signaling can lead to enhanced drought tolerance (Cao et al., 2017; Vaidya and Cutler, 2022) whereas downregulation of ABA signaling - when water sources are not limiting—can enhance crop yield (Miao et al., 2018) or serve to stimulate seed germination (Vaidya et al., 2021).

Altogether, these data have provided a deep understanding of the pathway that can be used to rationalize the effect of engineered mutations in *Arabidopsis* and crop plants or to design molecular sensors with tailored properties. In this work,

we have identified a single amino acid variation at the latch loop that classifies ABA receptors according to their oligomeric state. We have shown that a mutated dimeric PYL1 receptor from tomato (*Solanum lycopersicum*, SlPYL1) harboring this substitution renders to a receptor with enhanced ABA binding affinity and that inhibits PP2C in an ABA independent manner. The crystallographic analysis of the mutated protein has led to the identification of the basis of such behavior. Moreover, we employed a molecular dynamic based crystallographic refinement (Burnley et al., 2012) to show that the higher heterogeneity of the mutated latch loop eases the transition of the loop to a productive conformation for ABA binding. Moreover, the crystallographic analysis of the aforementioned mutant has led to the discovery of niacin as a new ABA antagonist, which blocks the latch loop by generating an intermediate and unproductive conformation of the receptor that prevents PP2C interaction.

MATERIALS AND METHODS

Mutagenesis of SlPYL1

The SlPYL1-E151D mutant was generated from pETM11-Sl08g076960wt (Gonzalez-Guzman et al., 2014) using the QuickChange Lightning Site-Directed Mutagenesis Kit (Agilent). Briefly, we performed a PCR using forward (GTATCATTGGCGGAGATCACCGGTTGAGGAATTACC) and reverse (GGTAATTCCTCAACCGGTGATCTCCGCCAATGATAC) primers carrying the desired mutation, followed by a DpnI digestion to restrict parental DNA. The digestion product was then transformed into *Escherichia coli* XL10 Gold cells. Positive colonies were sequenced to confirm the mutation.

Expression of SlPYL1 Wild Type and SlPYL1 E151D and Purification of His-Tagged Proteins

Escherichia coli BL21 (DE3) cells were transformed with the corresponding pETM11-SlPYL1 construct and grown at 37°C to an optical density at 600 nm of 0.7 in 1 L of 2-TY medium supplemented with 50 µg/ml kanamycin. Then, 0.3 mM isopropyl-β-D-thiogalactoside (IPTG) was added to the medium, and the cells were harvested after overnight incubation at 16°C. Pellets were resuspended in lysis buffer (30 mM Tris pH 7.5, 150 mM NaCl, 1 mM DTT) and lysed by sonication in a Branson sonifier. Clear lysates were obtained after centrifugation at 20,000 *g* for 40 min, and they were purified using a 1 ml nickel-nitrilotriacetic acid agarose column, previously equilibrated with 30 mM Tris pH 7.5, 150 mM NaCl, 1 mM DTT buffer. A washing step was performed using 30 mM Tris pH 7.5, 150 mM NaCl, 20 mM imidazole, 1 mM DTT, and the His tags were cleaved by tobacco etch virus protease. The proteins were eluted using 30 mM Tris pH 7.5, 150 mM NaCl, 40 mM imidazole, 1 mM DTT buffer. Eighteen and twenty-three milligram of SlPYL1 wild-type (WT) and SlPYL1 E151D were obtained and concentrated to 4.5 and 5.8 mg/ml, respectively. These samples were injected on a HiLoad Superdex 200 16/60 column (GE Healthcare)

previously equilibrated with 30 mM Tris pH 7.5, 150 mM NaCl, 1 mM DTT buffer. Fractions corresponding to dimeric SlPYL1 WT and E151D were pooled. Protein concentration was calculated using a Nanodrop spectrophotometer (Thermo Scientific) and their quality was checked by 12% SDS-PAGE (**Supplementary Figure 1**).

Crystallization, Data Collection, Structure Solution, and Refinement

The crystallization protocols for all structures were identical. In summary, crystals were obtained using vapor diffusion hanging drop technique using protein at 10 mg/ml concentration, 1.6 and 1.8M ammonium sulfate pH 7.0 as precipitant solution and a protein:precipitant ratio of 1:1 and 1:2. The crystals for SlPYL1 E151D:ABA and SlPYL1:niacin complexes were obtained using the same conditions by addition of 10 mM ABA (Biosynth) or niacin (Sigma Aldrich), respectively, to crystallization drops in a ratio 1:1:0.2 or 1:2:0.2. Crystals of SlPYL1 E151D:niacin complex grew from crystallization drops in a ratio 1:1 whereas crystals of SlPYL1 E151D apo protein grew from a ratio 1:2.

Crystals were cryoprotected in the crystallization solution containing 25–30% glycerol, mounted on a fiber loop and flash-frozen in liquid nitrogen. A complete diffraction data set was collected (see details in **Supplementary Table 1**). Diffraction data were processed with XDS (Kabsch, 2010) and merged with AIMLESS from the CCP4 package (Collaborative Computational Project, Number 4, 1994; Winn et al., 2011). We employed the crystal structures of SlPYL1 (PDB code 5MOA) and SlPYL1 in complex with ABA (PDB code 5MOB; Moreno-Alvero et al., 2017) to phase the diffraction data of crystal of the apo form of SlPYL1 E151D, the ABA and Niacin complexes and the SlPYL1 Niacin complex as the crystals were nearly isomorphous. Several cycles of restrained refinement with PHENIX (Adams et al., 2010) and iterative model building with COOT (Emsley and Cowtan, 2004) were required to obtain the final models where the waters were also modeled. Data collection, data processing and model refinement statistics are summarized in **Supplementary Table 1**. The stereochemistry of the models was verified with MolProbity and figures of molecular models were produced using PyMOL (The PyMOL Molecular Graphics System, Version 1.6.0.0 Schrödinger, LLC).

Modeling of Dynamics in SlPYL1 and SlPYL1 E151D Crystal Structures

To extract dynamical details from the X-ray data, the refined coordinates of native SlPYL1 were first re-refined using PHENIX (Adams et al., 2010). These coordinates and those of SlPYL1 E151D were used as input models for a time-averaged molecular dynamics refinement as implemented in the Phenix.ensemble-refinement routine, which was performed as described previously (Burnley et al., 2012). The ensemble refinement method performs molecular dynamic simulations to sample local atomic fluctuations of SlPYL1 in the crystal separating large-scale motions attributable to the rigid movement of secondary structures or lattice distortions. The method prevents over-fitting of the data by the restriction of the number of structures modeled

to 50 models. The calculated models for the apo forms of the SlPYL1 and SlPYL1 E151 shows improved fit to the X-ray data as R-free improved by 2.2 and 3.2 percentage points. Prior to the simulations, we approximated the large-scale disorder by an overall TLS model derived from the atomic B-factors of the refined single structure. The simulations were run at an effective temperature of 300 K for the protein atoms. The parameters p-tls, which is the percentage of atoms included in TLS-fitting, was optimized to 90%, 0.9, in a grid search to obtain the minimum R-free after the refinement.

PP2C Activity Assays

Phosphatase activity was measured using p-nitrophenyl phosphate (pNPP) as a substrate. The assays were performed in quadruplicates in a 100 µl solution containing 25 mM Tris–HCl pH 7.5, 75 mM NaCl, 0.5 mM DTT, 1 mM MnCl2, and 25 mM pNPP. The assays included molar relations of phosphatase:receptor of 1:2 or 1:4 (0.5 µM ΔN-HAB1 and 1 or 2 µM SlPYL1 WT or mutant), and the indicated concentration of ABA or nicotinic acid. We chose the PP2C from *Arabidopsis* HAB1 because successful procedures to determine the half maximal inhibitory concentrations (IC_{50}) as ABA receptors are able to inhibit PP2Cs from different plant species and successful protocols using HAB1 have been previously reported (Moreno-Alvero et al., 2017; Lozano-Juste et al., 2021). The activity was recorded with a FLUOstar Omega plate reader at 405 nm every 5 min over 20 min, and the activity obtained after 20 min was indicated in the graphs.

Thermal Shift Assay

Label-free thermal shift assays with SlPYL1 WT and SlPYL1 E151D were performed using Tycho NT. 6 (NanoTemper Technologies, Germany). Assays were performed at 6 µM protein in 30 mM Tris–HCl, pH 7.5, 150 mM NaCl, 1 mM DTT and run in quadruplicates in Tycho NT.6 capillaries (Cat no. TY-C001; NanoTemper Technologies, Germany), using final ABA concentrations of 0.067, 0.25, 1, and 4 mM. Intrinsic fluorescence was recorded at 330 and 350 nm while heating the sample from 35 to 95°C at a rate of 30°C/min. Fluorescence ratio (350/330 nm) and inflection temperature (Ti) were calculated by Tycho NT. 6.

Size Exclusion Chromatography

Refractive index measurements were carried out at different protein concentrations. 50 µl of protein solutions with 60, 6, or 0.6 µg of SlPYL1 WT or mutant in 30 mM Tris–HCl, pH 7.5, 150 mM NaCl, 1 mM DTT buffer were loaded on a Superdex 30 Increase 3.2/300 column (GE Healthcare). The refractive index of the samples was recorded over time using a Shimadzu LC-20AD Prominence HPLC Pump (Shimadzu Scientific Instruments).

Plants

Approximately 25 seeds (two replicates per experiment) of *Arabidopsis thaliana* Col-0 seeds were sown on 24-multiwell plates containing MS medium, which was mock- or ABA-supplemented at 0.125 or 0.250 µM concentration. Increasing concentrations (from 0.5 to 10 µM) of nicotinic acid were added
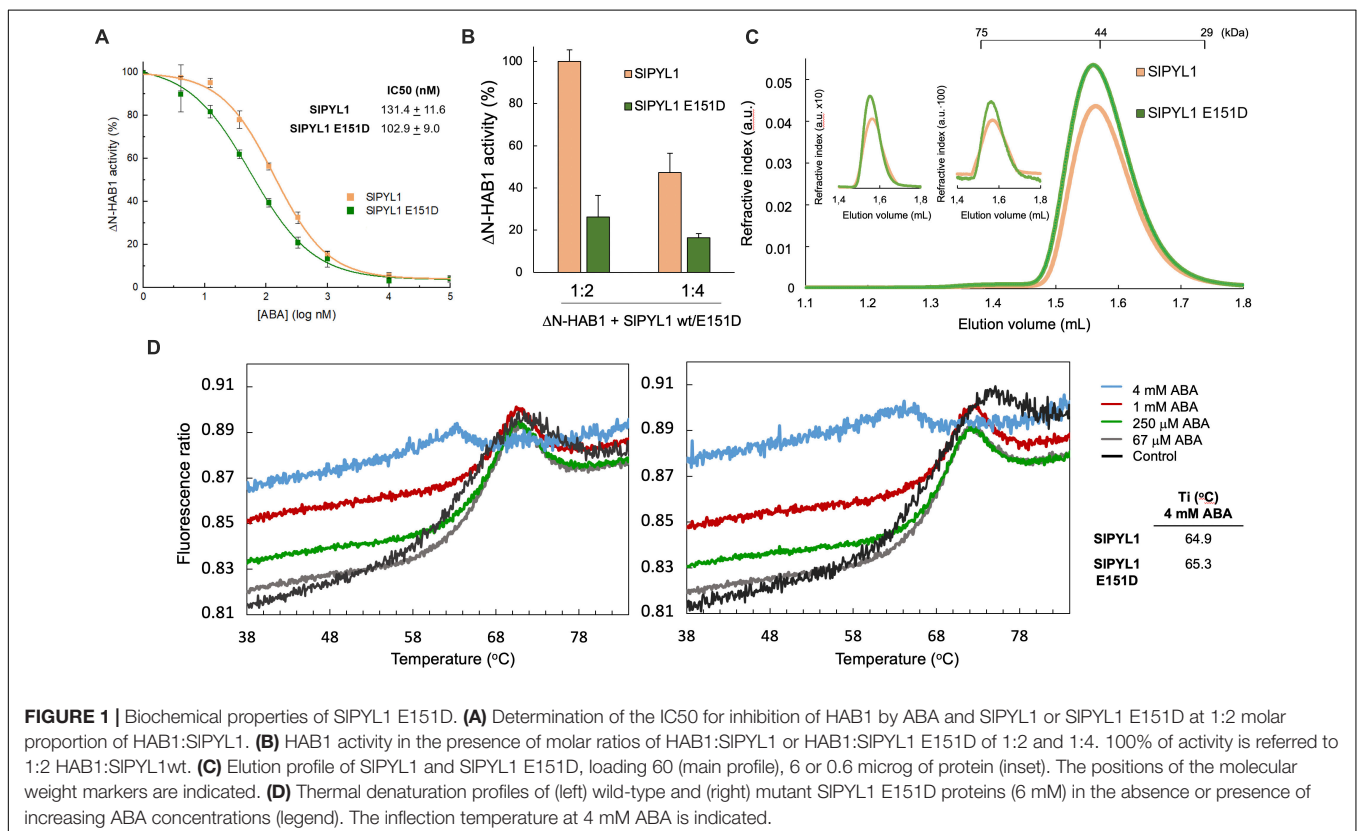

FIGURE 1 | Biochemical properties of SlPYL1 E151D. **(A)** Determination of the IC50 for inhibition of HAB1 by ABA and SlPYL1 or SlPYL1 E151D at 1:2 molar proportion of HAB1:SlPYL1. **(B)** HAB1 activity in the presence of molar ratios of HAB1:SlPYL1 or HAB1:SlPYL1 E151D of 1:2 and 1:4. 100% of activity is referred to 1:2 HAB1:SlPYL1wt. **(C)** Elution profile of SlPYL1 and SlPYL1 E151D, loading 60 (main profile), 6 or 0.6 microg of protein (inset). The positions of the molecular weight markers are indicated. **(D)** Thermal denaturation profiles of (left) wild-type and (right) mutant SlPYL1 E151D proteins (6 mM) in the absence or presence of increasing ABA concentrations (legend). The inflection temperature at 4 mM ABA is indicated.

in mock- or ABA-supplemented plates. Seedling establishment and early seedling growth was scored at 5 and 9 days after sowing. Root length was scored at 5 days after sowing.

RESULTS

Glutamic to Aspartic Acid Substitution at Latch Loop of SlPYL1 Enhances Receptor Ability to Inhibit the Phosphatase HAB1

The comparison of the available structures of PYR/PYL receptors showed that latch closure is a prerequisite for ABA binding, in turn for gate closure and for the subsequent phosphatase inhibition (Melcher et al., 2009; Moreno-Alvero et al., 2017; Rodriguez et al., 2019). This conformational rearrangement involves the concerted movement of a conserved Glu/Asp-His-Arg motif at the latch to make the binding site accessible to ABA and to define its binding pocket. Interestingly, while the His-Arg sequence is conserved among the whole family of PYR/PYL receptors, the substitution Glu to Asp differentiates the group of the dimeric receptors from the monomeric ones (**Supplementary Figure 2**). Accordingly, the amino acid sequence alignment of tomato receptors showed that the corresponding residue in monomeric PYLs of tomato is also Asp (Gonzalez-Guzman et al., 2014). To investigate the relevance of this amino acid on the receptor function, we generated a mutated version of SlPYL1 harboring a Glu 151 to Asp substitution (SlPYL1 E151D) at the latch and studied its biochemical properties and compared them with those of the native protein.

We first analyzed the PP2C inhibitory activity of the WT and mutated receptor *in vitro*. Our data showed that IC_{50} value of ABA using SlPYL1 E151D was 103 ± 9 nM which is significantly smaller than that calculated using WT protein 131 ± 11 nM (**Figure 1A**). In addition, we examined the inhibitory effect of both WT and mutant receptors in absence of ABA with HAB1:receptor ratios of 1:2 and 1:4. Interestingly, we observed a higher reduction in phosphatase activity using SlPYL1 E151D when compared to SlPYL1 (**Figure 1B**).

The enhanced activity of the mutated receptor suggests that the Asp 151 residue might be involved in the dissociation of the dimeric structure despite it is located apart from the oligomerization interface. Hence, we investigated the oligomeric state of the WT and SlPYL1 E151D mutant protein by Size-Exclusion Chromatography (SEC). We performed a series of SEC experiments varying the protein concentration injected from 48 to 0.48 μM revealing that the mutation does not affect the oligomeric state of the apo form of both proteins as they form dimers in solution at all the assayed concentrations (**Figure 1C**).
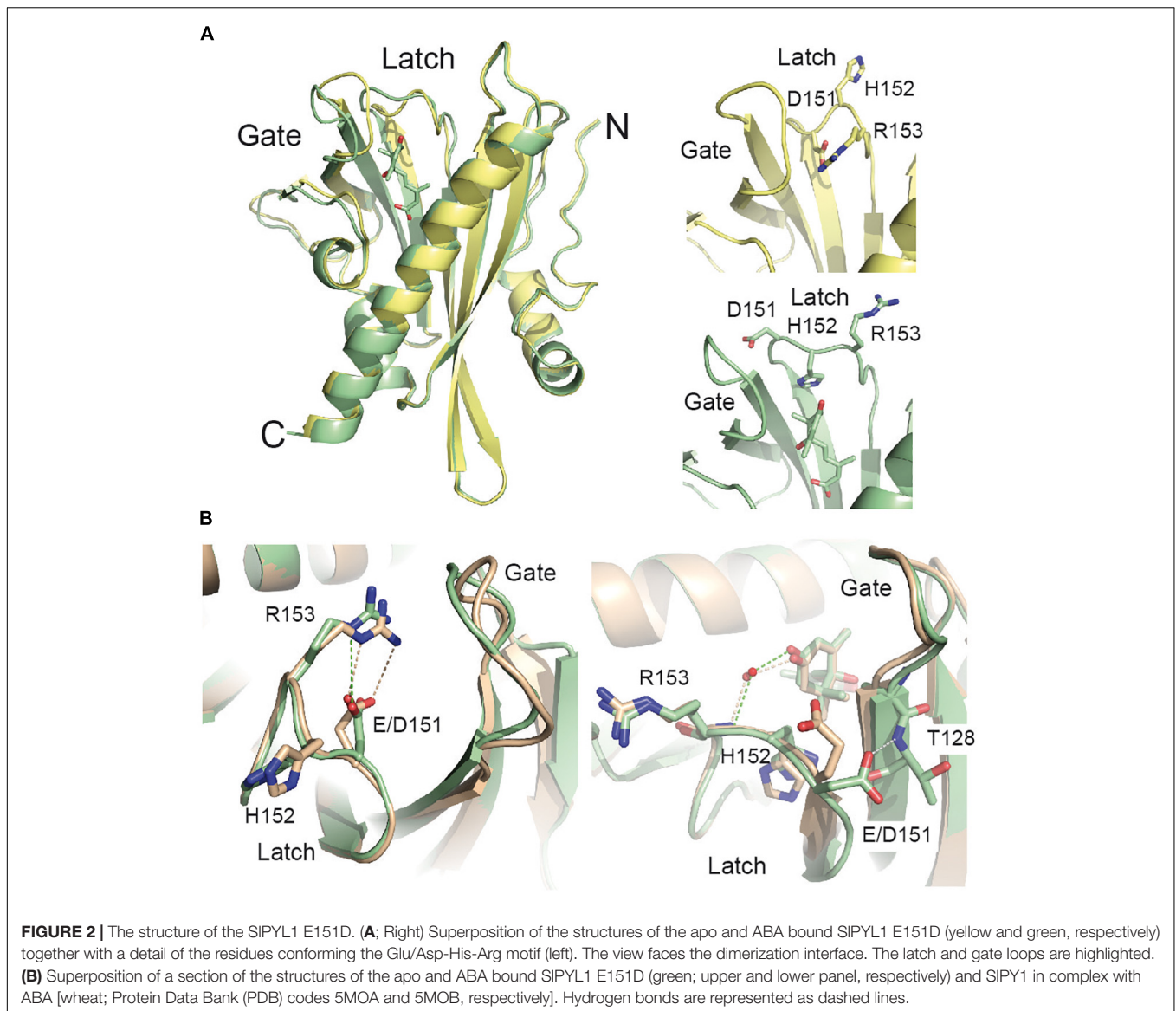
Protein stability and biological activity are usually coupled when comparing native and mutated enzymatic systems (Rodriguez-Larrea et al., 2010; Mesa-Torres et al., 2014). We therefore studied the effect of the amino acid substitution on the stability of the apo and ABA bound forms of the protein using a label-free thermal shift assay (nano-Differential Scanning Fluorimetry, nanoDSF, Tycho NT. 6, NanoTemper Technologies,

Germany). Using this technique, the intrinsic fluorescence of the receptors without or at increased concentrations of free ABA were recorded at 330 and 350 nm while heating the sample from 35 to 95°C. The variation of the ratio of fluorescence (350/330 nm) versus the Ti usually follows a sigmoidal curve which is related to the cooperative thermal denaturation of the protein. The analysis of this curve provides an inflection Ti that reports on the stability of the protein or on the formation of a complex with a ligand if a binding event occurs. However, in absence of ABA, complex denaturation profiles of both WT and SlPYL1 E151D proteins were observed. Interestingly, this complexity is reduced at increased ABA concentrations to a point at 4 mM ABA in which a simple sigmoidal transition is observed. As the dimeric receptors tend to dissociate upon ABA binding (Dupeux et al., 2011b), this suggests that the transition from native to unfolded protein at low ABA concentration occurs through two independent events, the dissociation of the dimeric species and the denaturation of the monomeric protein (**Figure 1D**). The complex nature of the denaturation profiles obtained without or at low concentrations of ABA hinders the calculation of Ti, however, SlPYL1 E151D protein displays larger Ti than the WT in the presence of 4 mM ABA, showing that the Asp side chain is involved in the stabilization of the protein.

The Structure SlPYL1 E151D Provides the Basis for Its Enhanced Abscisic Acid Dependent Activity

The structure of the SlPYL1 and other ABA receptors revealed that Glu 151 is located in the interface of the gate and latch loops (Moreno-Alvero et al., 2017), hence, we reasoned that E151D mutation might be affecting the conformational rearrangements leading the receptor activation. To investigate this question and to determine the structural basis of the effect of E151D mutation on ABA perception, we underwent crystallographic studies (**Figure 2**, **Supplementary Table 1**, and section "Materials and Methods"). The overall structures of the apo and ABA bound forms of SlPYL1 E151D are nearly identical to that of WT protein with the differences confined to the area of the mutated residue. The Root Mean Squared Deviation for 191 Calpha atoms are of 0.33A and 0.20A for the apo and ABA bound forms, respectively.

Previous crystallographic data on native SlPYL1 showed that the transition from the closed to the open conformation of the latch included a large rearrangement of Glu 151 (Moreno-Alvero et al., 2017). In this structure, the apo form showed that Glu 151 and Arg 153 established a salt bridge that blocked the ABA binding pocket while, in the ABA bound form, Glu 151 pointed to the solvent enabling ABA binding and subsequent gate closure (Moreno-Alvero et al., 2017; **Figure 2B**). Instead, the analysis of the apo form of SlPYL1 E151D structure revealed that the shortening of the Glu 151 side chain to Asp disrupts the geometry of the salt bridge with Arg 153, weakening this interaction and favoring the transition to an active receptor. In addition, the ABA bound structure shows that Asp 151 carboxylate is hydrogen bonded to the NH of Thr 128 at the C-terminal end of the gate loop. This hydrogen bond is not present in the native structure and might account for the increased stability of the mutant

FIGURE 2 | The structure of the SlPYL1 E151D. (**A**; Right) Superposition of the structures of the apo and ABA bound SlPYL1 E151D (yellow and green, respectively) together with a detail of the residues conforming the Glu/Asp-His-Arg motif (left). The view faces the dimerization interface. The latch and gate loops are highlighted. (**B**) Superposition of a section of the structures of the apo and ABA bound SlPYL1 E151D (green; upper and lower panel, respectively) and SlPY1 in complex with ABA [wheat; Protein Data Bank (PDB) codes 5MOA and 5MOB, respectively]. Hydrogen bonds are represented as dashed lines.

protein in complex with ABA with respect to that observed for the native protein (**Figure 1D**). Interestingly, the ABA bound forms of the receptors display a wider aperture at the binding site that the apo forms, this feature is also observed for citrus PYL1 and other crop receptors (Moreno-Alvero et al., 2017; **Figure 2**).

These results highlight the importance of the dynamical aspects on ABA perception and pathway activation. Thus, to gain insights into these issues, we performed a molecular dynamic based ensemble refinement restrained by crystallographic data (Adams et al., 2010; **Figure 3A**, **Supplementary Table 1**, and section "Materials and Methods"). This protocol provides a realistic picture of the molecular fluctuations of a protein within a crystal by assembling multiple models of its structure (Burnley et al., 2012). Hence, in combination with the traditional crystallographic refinement, the protocol enables the investigation of the consequences of the E151D mutation on the dynamic properties of gate and latch loops.

The analysis of the superimposition and the root mean square deviation (RMSD) per residue along the 50 models that result from the ensemble refinement illustrates the dynamics in SlPYL1 and SlPYL1 E151D crystal structures. The increased RMSD of the gate and latch loops in the apo forms of SlPYL1 E151D with respect to those of SlPYL1 (**Figures 3A,B**) shows that mutated protein displays higher conformational heterogeneity at these loops. In addition, it reveals that the native structure displays a well-ordered salt bridge between Glu 151 and Arg 153 blocking ABA binding while the mutant protein displays high flexibility for Asp 151 and Arg 153 (**Figure 3C**), thus suggesting that the Asp151 mutation might ease the large transition – of the latch for a productive ABA recognition. Interestingly, there is a reduction of the RMSD as a result of ABA binding, that indicates that hormone binding stabilizes the entire gate loop despite it does not make any direct contact with ABA (**Figures 3A,B**). However, the degree of such stabilization is higher in the mutant protein.
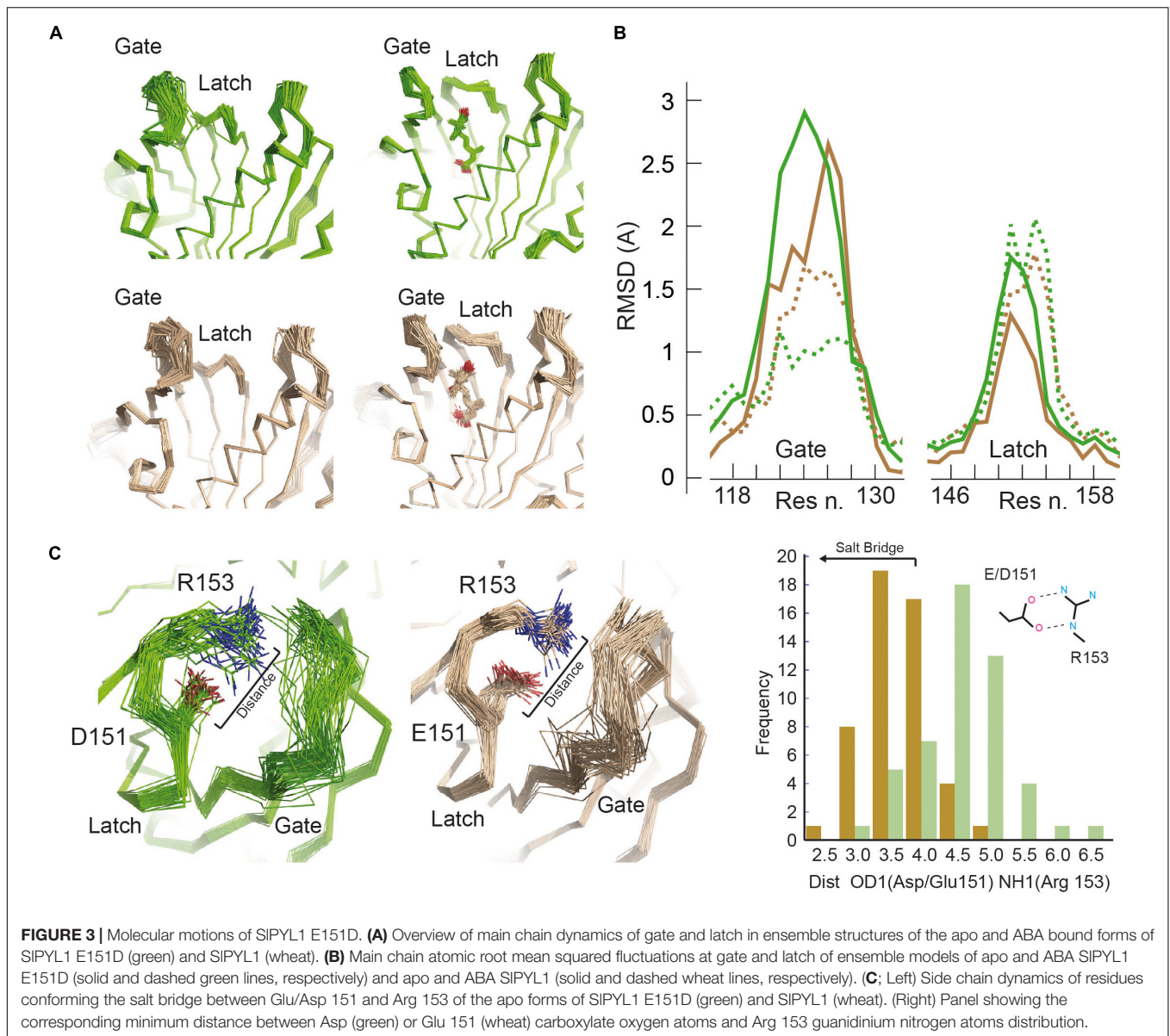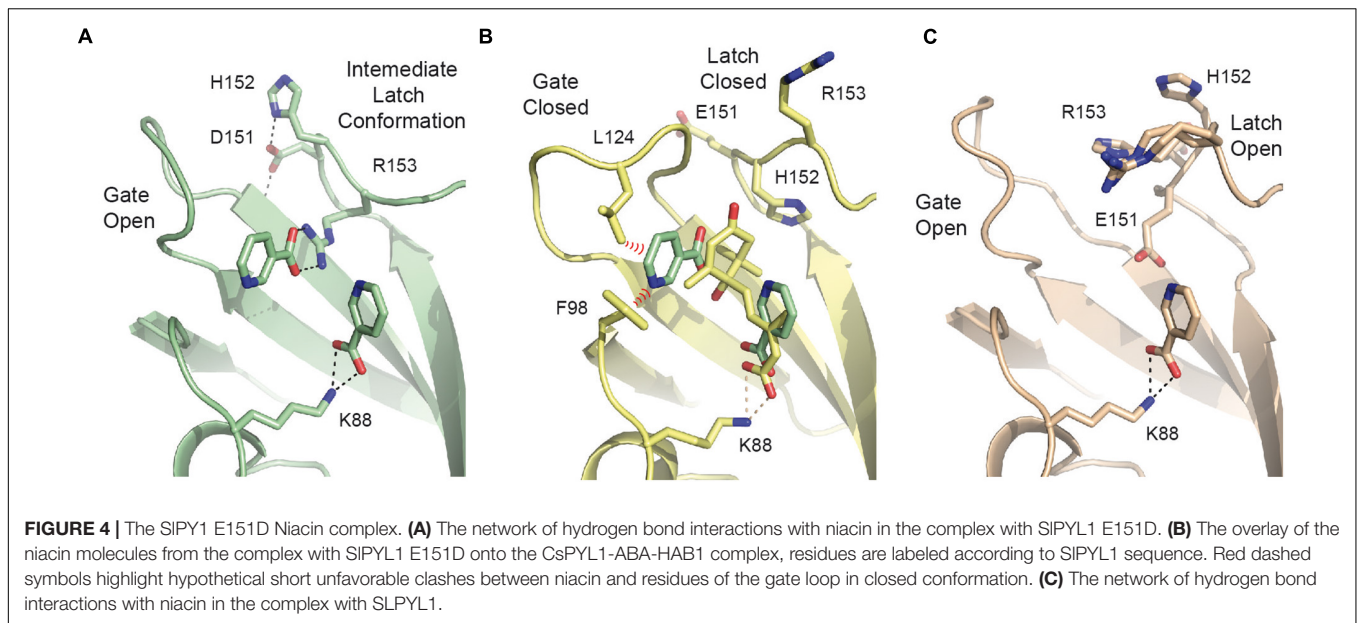
FIGURE 3 | Molecular motions of SlPYL1 E151D. **(A)** Overview of main chain dynamics of gate and latch in ensemble structures of the apo and ABA bound forms of SlPYL1 E151D (green) and SlPYL1 (wheat). **(B)** Main chain atomic root mean squared fluctuations at gate and latch of ensemble models of apo and ABA SlPYL1 E151D (solid and dashed green lines, respectively) and apo and ABA SlPYL1 (solid and dashed wheat lines, respectively). **(C**; Left) Side chain dynamics of residues conforming the salt bridge between Glu/Asp 151 and Arg 153 of the apo forms of SlPYL1 E151D (green) and SlPYL1 (wheat). (Right) Panel showing the corresponding minimum distance between Asp (green) or Glu 151 (wheat) carboxylate oxygen atoms and Arg 153 guanidinium nitrogen atoms distribution.

The Crystallographic Studies on SlPYL1 E151D Unravels Niacin as Antagonist Molecule of Abscisic Acid Activity

The crystals of SlPYL1 E151D grown at 1:1 protein:precipitant ratio without ABA displayed two residual blobs of density at the ABA site that we interpreted chemically as two nicotinic acid molecules, also called niacin, product of bacterial expression (**Figure 4A** and **Supplementary Figure 3A**). Interestingly, the ligand molecules mimic the position and some of the interactions observed for ABA in the CsPYL1-ABA-HAB1 ternary complex (Moreno-Alvero et al., 2017; **Figure 4B**). The first site is interacting with the α amino group NZ of Lys 88 and mimics the carboxylate moiety of ABA, while the second interacts with the indole moiety Arg 153. In this situation, Asp 151 adopts the ABA bound conformation while His 152 points toward solvent

as it is observed in the apo crystal form. Thus, this conformation represents an intermediate between that observed for the apo and ABA bound structures. Remarkably, Asp 151 is not interacting with niacin, rather it is involved in the stabilization of the conformation of the loop via side chain hydrogen bonds to NH of Thr 128 and NE of His 152.

Attempts to reproduce the crystallization conditions yielding to a niacin complex with the WT protein were unsuccessful, hence we grew crystals of SlPYL1 with a solution containing 1.0 mM niacin and solved the crystal structure to investigate the role of Glu to Asp 151 substitution in the loop conformation. Our data show that the WT protein interacts solely with a molecule of niacin at ABA carboxylate site which is interacting with NZ of Lys 88 (**Figure 4C** and **Supplementary Figure 3B**). These data suggest that formation of the complex with two molecules of niacin would be a consequence of the higher

FIGURE 4 | The SlPY1 E151D Niacin complex. **(A)** The network of hydrogen bond interactions with niacin in the complex with SlPYL1 E151D. **(B)** The overlay of the niacin molecules from the complex with SlPYL1 E151D onto the CsPYL1-ABA-HAB1 complex, residues are labeled according to SlPYL1 sequence. Red dashed symbols highlight hypothetical short unfavorable clashes between niacin and residues of the gate loop in closed conformation. **(C)** The network of hydrogen bond interactions with niacin in the complex with SLPYL1.

flexibility of the mutated latch loop with respect to that of the WT structure (**Figure 3**).

Our crystallographic data shows that the interaction of niacin at the ABA carboxylate site both in the native and in the mutant protein, would compete with ABA for the binding site and might provide the basis for a universal ABA antagonist molecule. Moreover, niacin hinders gate and latch closure of the SlPYL1 E151D mutant protein (**Figure 4B**), hence, we reasoned that as monomeric receptors naturally harbor Asp at the Glu/Asp-His-Arg latch motif, niacin may represent an effective ABA antagonist of this family of receptors.
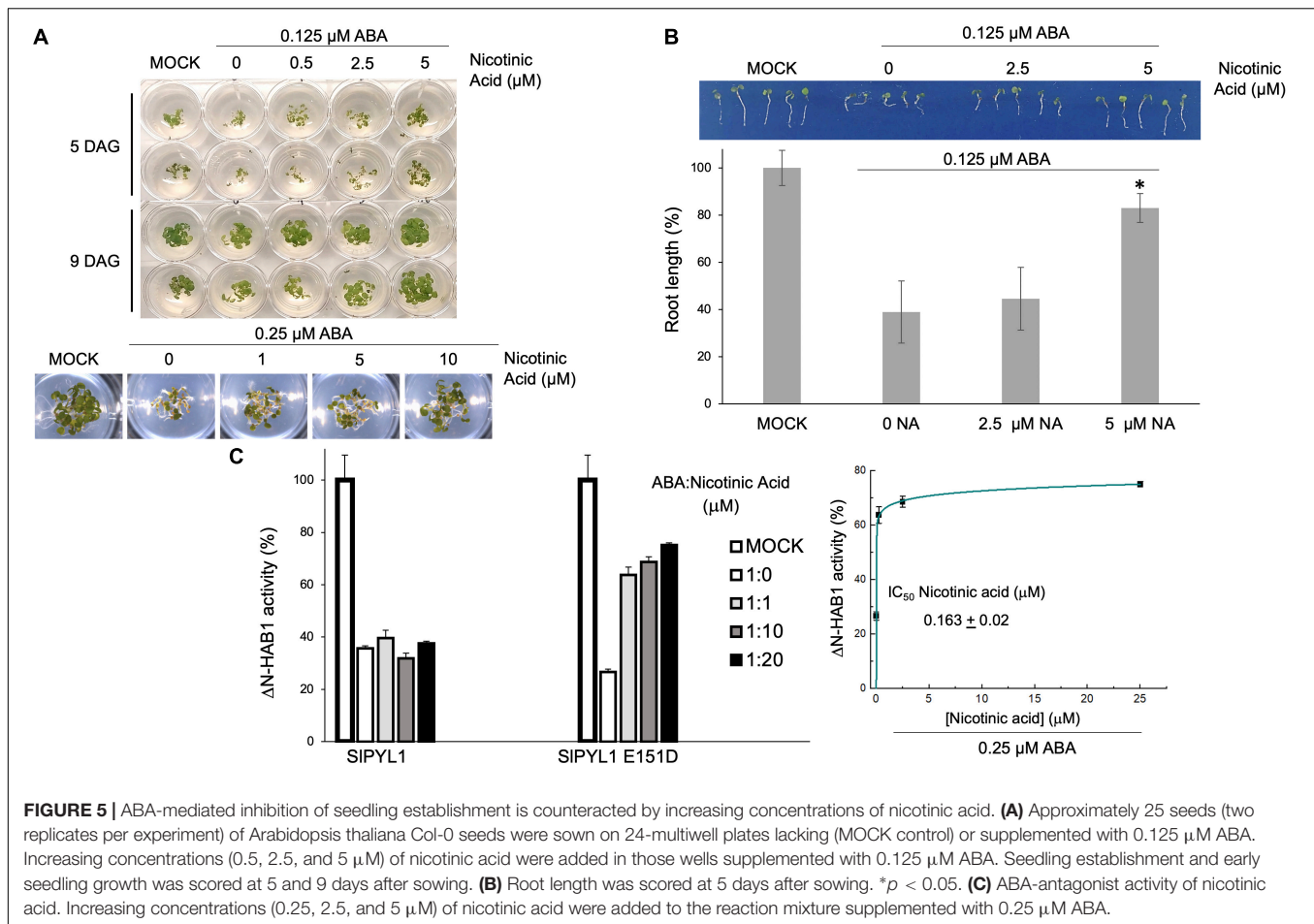
To test whether niacin can antagonize the effects of ABA *in vivo*, we performed *Arabidopsis* seed germination/seedling establishment assays. ABA inhibits seedling establishment and early seedling growth; therefore, we analyzed whether different concentrations of niacin coappled with ABA could restore seed establishment of *Arabidopsis*. **Figure 5A** shows that increasing concentrations of niacin relieve the inhibition of seedling establishment and early seedling growth when coappled with ABA. This was apparent at concentrations between 20- and 40-fold higher than that of ABA. In particular, **Figure 5B** shows that root length was restored when 5 μM Niacin was coappled with 0.125 μM ABA. Thus, this revealed niacin as an ABA antagonist *in vivo*.

To investigate whether the antagonistic effect of niacin was mediated by the binding of the compound to the carboxylate site or to the latch loop of the receptor, we examined the ability of niacin to reverse ABA mediated PP2C inhibition using WT and mutant SlPYL1. Our data showed that niacin produces a clear protective effect against ABA enabling a recovery of HAB1 activity even at equimolecular concentrations of ABA and niacin in the presence of SlPYL1 E151D (**Figure 5C**). Conversely, no effect was observed when assaying the WT protein. This shows that antagonistic effect of niacin requires the presence of Asp 151 at the latch loop.

DISCUSSION

The design of an ABA receptor of an economically valuable crop specie to be ABA independent or to display an exacerbated response to ABA represents a biotechnological opportunity to produce genetically modified plant varieties with enhanced performance in drought stress situations. Tomato ABA receptors have been characterized (Gonzalez-Guzman et al., 2014) and, as those from *Arabidopsis* (Dupeux et al., 2011b), they are classified as monomeric or dimeric according to their oligomeric state. As the monomeric receptors display higher ABA binding affinities than the dimeric ones and are able to form low affinity ABA independent complexes with PP2Cs, it is sensible to identify those residues that are conserved among the members of each class and analyze if they are involved in the receptor activity. Indeed, previous work led to the identification of a single point mutation in the dimerization interface of PYR1 ABA receptor that enhances the ABA binding affinity one order of magnitude despite this mutation is localized apart from the hormone binding site (Dupeux et al., 2011b). To investigate this issue, we have selected as a model SlPYL1 which is a tomato dimeric receptor with known structure (Moreno-Alvero et al., 2017) that is involved in the induction of stress responsive genes and drought tolerance (Gonzalez-Guzman et al., 2014).

The sequence analysis of SlPYL1 and those ABA receptors from *Arabidopsis* led to the identification of Glu 151 which is a unique position in the alignment as it is conserved among the dimeric receptors and that is replaced by Asp in the monomeric ones. Interestingly, the analysis of the structure of SlPYL1 shows that Glu 151 is not involved in the dimerization interface or interacting with ABA; rather, it is located in the latch loop (**Figure 2**). Hence, to investigate the effect of this substitution, we prepared a SlPYL1 E151D protein and performed biochemical and structural studies. As expected, we did not observe any effect on the oligomeric properties of the mutated receptor (**Figure 1C**).
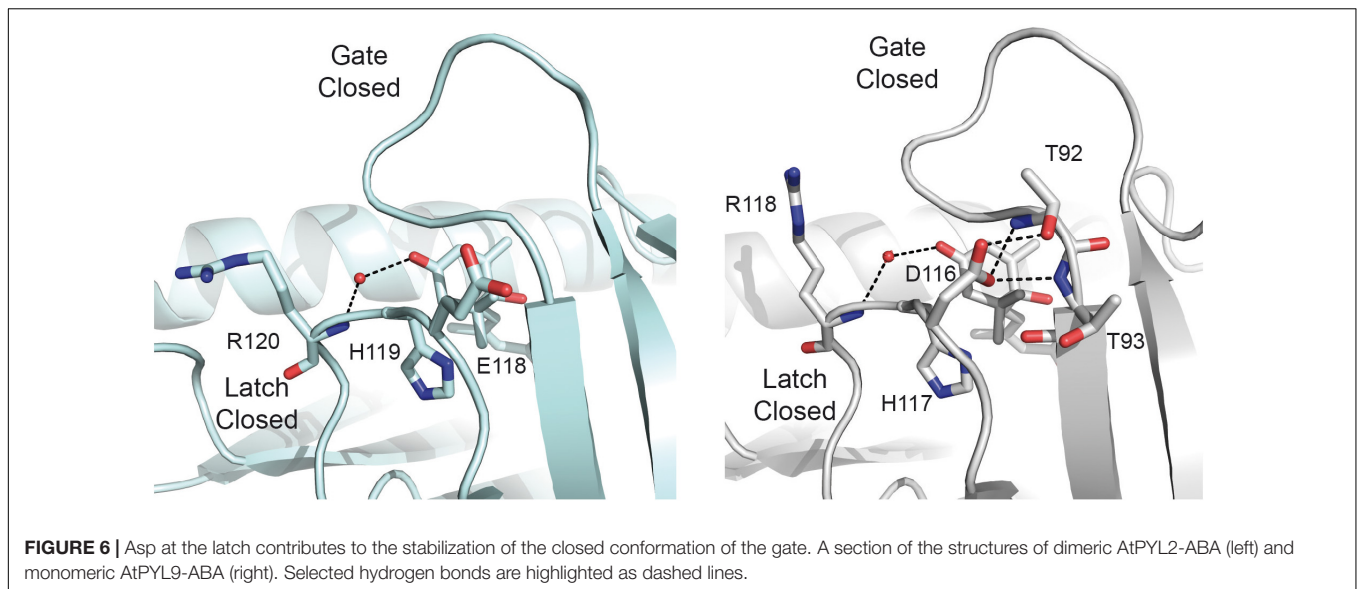
FIGURE 5 | ABA-mediated inhibition of seedling establishment is counteracted by increasing concentrations of nicotinic acid. **(A)** Approximately 25 seeds (two replicates per experiment) of Arabidopsis thaliana Col-0 seeds were sown on 24-multiwell plates lacking (MOCK control) or supplemented with 0.125 μM ABA. Increasing concentrations (0.5, 2.5, and 5 μM) of nicotinic acid were added in those wells supplemented with 0.125 μM ABA. Seedling establishment and early seedling growth was scored at 5 and 9 days after sowing. **(B)** Root length was scored at 5 days after sowing. *$p < 0.05$. **(C)** ABA-antagonist activity of nicotinic acid. Increasing concentrations (0.25, 2.5, and 5 μM) of nicotinic acid were added to the reaction mixture supplemented with 0.25 μM ABA.

However, our data showed that Glu 151 to Asp enhances the ability of SlPYL1 to inhibit the phosphatase HAB1 in an ABA dependent and independent manner when compared with the WT protein (**Figures 1A,B**). This suggests that the mutation is stimulating the conformational rearrangements of the latch loop required for ABA recognition. Indeed, our structural studies corroborated that the mutated latch loop displays higher heterogeneity than the WT, due to the disruption of a salt bridge between Glu 151 and Arg 153 which is capping the entrance to the ABA binding site in the apo form of the receptor (**Figures 2**, **3**). In addition, the analysis of the ABA bound structures of SlPYL1 E151D and SlPYL1 showed that Glu 151 replacement to Asp leads to the formation of an additional hydrogen bond linking Asp 151 at the latch and Thr 110 at the gate (**Figure 2B**). This would explain our data showing that the ABA bound form of SlPYL1 E151D displays increased stability as compared with that of SlPYL1 (**Figure 1D**).

Altogether, the joined analysis of the biochemical and crystallographic data suggests that the weakening of the capping salt bridge in the apo structure of the mutant protein together with the formation of an additional hydrogen bond linking Asp 151 and Thr 110 in the ABA bound structure shift the equilibrium from the apo form to the active form of the mutant receptor. This would not be sufficient to dissociate the dimer, but would

enhance the ABA dependent and independent binding activity. This suggests that the presence of Glu side chain in the Glu/Asp-His-Arg latch motive of dimeric receptors contributes to increase their ABA dependent activity with respect to the monomeric ones in addition to the unfavorable contribution derived from dimer dissociation (Dupeux et al., 2011b; Okamoto et al., 2013). To explore further whether the Glu to Asp substitution stabilizes a productive gate conformation, we analyzed the crystal structures of the ABA forms of the *Arabidopsis* receptors available in the literature. These structures display a productive gate conformation despite they are not in complex with a PP2C. Our study indicates that while monomeric AtPYL9 shows that Asp 117 is hydrogen bonded to Thr 92 and Thr 93 NH at the gate (Zhang et al., 2013), the Glu 188 side chain of the dimeric AtPYL2 is fully exposed to the solvent (Yin et al., 2009; **Figure 6**). Identical situation can be observed for monomeric AtPYL10 (PDB code: 3R6P) and for dimeric AtPYR1 (Nishimura et al., 2009), AtPYL1 (Miyazono et al., 2009), and AtPYL3 (Zhang et al., 2012), respectively. This data indicates that as observed in SlPYL1 E151D, the Asp side chain at the latch loop contributes to the stabilization of a productive conformation of the monomeric receptors.

SlPYL1 E151D displays an ABA independent phosphatase inhibitory activity (**Figure 1C**). Other approaches have been

FIGURE 6 | Asp at the latch contributes to the stabilization of the closed conformation of the gate. A section of the structures of dimeric AtPYL2-ABA (left) and monomeric AtPYL9-ABA (right). Selected hydrogen bonds are highlighted as dashed lines.

employed to identify specific substitutions to generate activated dimeric receptors in the absence of ABA. These include the mutation of PYL4 residues located in the interface between receptor and PP2C (Pizzio et al., 2013) or to perform random site-saturation mutagenesis of AtPYR1 receptor (Mosquna et al., 2011). The latter was successful for the identification of combinations of three- and four-point mutations that led to a constitutively active receptor adopting gate closed conformation. Interestingly, while these mutations were located at the dimerization interface, Glu 151 to Asp mutation represents an additional hot spot as it is located in the latch loop. Altogether, our work illustrates that the high-resolution structural information of PYR/PYL ABA receptors may be employed to rationally design a receptor with ABA dependent and independent enhanced properties.

The crystallographic analysis of SlPYL1 E151D led to the identification of niacin as an ABA antagonist molecule. Our data showed that two molecules of niacin were blocking the ABA binding site and stabilizing the latch loop in an unproductive conformation that would hinder gate closure. Remarkably, the purification protocol of SlPYL1 E151D (section "Materials and Methods") includes a buffer exchange step after Ni^{2+} affinity chromatography and a slow size exclusion chromatographic step on a 120 ml Superdex 200 column, thus indicating that the binding of niacin to SlPYL1 E151D displays slow kinetics. However, we produced an estimation of niacin potence as antagonist using HAB1 phosphatase recovery assay in the presence of 0.25 μM ABA. Our data shows that niacin reverses ABA-mediated HAB1 phosphatase inhibition using SlPYL1 E151D in the sub-micromolar range (**Figure 5C**). However, in comparison to other antagonists, niacin is 100-fold less potent than the best-known ABA antagonist molecule (Vaidya et al., 2021).

The effect of niacin *in vivo* was investigated using seed germination/seedling establishment assays. These experiments rely on the inhibitory effect of ABA on germination of *Arabidopsis*

seeds. Our data showed that although niacin reverted this inhibitory effect when coapplied with ABA, this activity required a high molar excess of compound to ABA to be observed (**Figure 4**). However, Niacin is a precursor of Nicotinamide adenine dinucleotide (NAD), which has been documented in plants to function in response to environmental stresses including pathogen infections (Alferez et al., 2018). Plant levels of NAD and its precursors are in the millimolar or sub-millimolar range (Zhang and Mou, 2009); therefore, the observed effect with micromolar niacin concentration might be physiological.

To investigate the molecular basis of niacin activity, we performed SlPYL1 and SlPYL1 E151D mediated PP2C inhibition assays. Our data showed that while niacin is able to reverse ABA mediated HAB1 phosphatase inhibition using SlPYL1 E151D, it does not display any effect in presence of WT protein. This indicates that despite Asp 151 is not directly interacting with niacin it is required for its antagonistic activity. Consequently, the observed *in vivo* activity might be mediated by monomeric receptors as they harbor Asp at the Glu/Asp-His-Arg latch motif. This is relevant as niacin might be useful for dissecting physiological roles of monomeric ABA receptors *in vivo*. In addition, as intracellular basal ABA levels are sufficient to mediate ABA response of monomeric receptors (Tischer et al., 2017), it is tempted to speculate that the physiological basal of niacin might be modulating the ABA affinity of those monomeric receptors and consequently their activity in response to stress situations. Besides, niacin might constitute a natural precursor of new environmentally friendly family of agrochemical compounds application to fight against drought stress.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and

accession number(s) can be found below: http://www.wwpdb.org/, 7Z1P; http://www.wwpdb.org/, 7Z1Q; http://www.wwpdb.org/, 7Z1R; and http://www.wwpdb.org/, 7Z1S.

AUTHOR CONTRIBUTIONS

AA: conceptualization, formal analysis, investigation, and writing–original draft preparation. LI and AA: crystallization and structural work. MR-M, MD-M, and JB: protein expression and purification. MR-M: activity and characterization assays. JO-C, AC, JL-J, and PR: plant assays. All authors contributed to the manuscript version and editing and approved the submitted version of the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.884029/full#supplementary-material

REFERENCES

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution title. *Acta Crystallogr. D Biol. Crystallogr.* 66, 213–221. doi: 10.1107/s0907444909052925

Alferez, F. M., Gerberich, K. M., Li, J. L., Zhang, Y., Graham, J. H., and Mou, Z. (2018). Exogenous Nicotinamide Adenine Dinucleotide Induces Resistance to Citrus Canker in Citrus. *Front. Plant Sci.* 9:1472. doi: 10.3389/fpls.2018.01472

Antoni, R., Gonzalez-Guzman, M., Rodriguez, L., Rodrigues, A., Pizzio, G. A., and Rodriguez, P. L. (2012). Selective inhibition of clade A phosphatases type 2C by PYR/PYL/RCAR abscisic acid receptors. *Plant Physiol.* 158, 970–980. doi: 10.1104/pp.111.188623

Burnley, B. T., Afonine, P. V., Adams, P. D., and Gros, P. (2012). Modelling dynamics in protein crystal structures by ensemble refinement. *Elife* 1:e00311. doi: 10.7554/eLife.00311

Cao, M., Liu, X., Zhang, Y., Xue, X., Zhou, X. E., Melcher, K., et al. (2013). An ABA-mimicking ligand that reduces water loss and promotes drought resistance in plants. *Cell Res.* 23, 1043–1054. doi: 10.1038/cr.2013.95

Cao, M. J., Zhang, Y. L., Liu, X., Huang, H., Zhou, X. E., Wang, W. L., et al. (2017). Combining chemical and genetic approaches to increase drought resistance in plants. *Nat. Commun.* 8:1183. doi: 10.1038/s41467-017-01239-3

Chen, P., Sun, Y. F., Kai, W. B., Liang, B., Zhang, Y. S., Zhai, X. W., et al. (2016). Interactions of ABA signaling core components (SlPYLs, SlPP2Cs, and SlSnRK2s) in tomato (Solanum lycopersicon). *J. Plant Physiol.* 205, 67–74. doi: 10.1016/j.jplph.2016.07.016

Cutler, S. R., Rodriguez, P. L., Finkelstein, R. R., and Abrams, S. R. (2010). Abscisic acid: emergence of a core signaling network. *Annu. Rev. Plant Biol.* 61, 651–679. doi: 10.1146/annurev-arplant-042809-112122

Dupeux, F., Antoni, R., Betz, K., Santiago, J., Gonzalez-Guzman, M., Rodriguez, L., et al. (2011a). Modulation of abscisic acid signaling in vivo by an engineered receptor-insensitive protein phosphatase type 2C allele. *Plant Physiol.* 156, 106–116. doi: 10.1104/pp.110.170894

Dupeux, F., Santiago, J., Betz, K., Twycross, J., Park, S. Y., Rodriguez, L., et al. (2011b). A thermodynamic switch modulates abscisic acid receptor sensitivity. *EMBO J.* 30, 4171–4184. doi: 10.1038/emboj.2011.294

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics title. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132. doi: 10.1107/S0907444904019158

Gonzalez-Guzman, M., Pizzio, G. A., Antoni, R., Vera-Sirera, F., Merilo, E., Bassel, G. W., et al. (2012). Arabidopsis PYR/PYL/RCAR receptors play a major role in quantitative regulation of stomatal aperture and transcriptional response to abscisic acid. *Plant Cell* 24, 2483–2496. doi: 10.1105/tpc.112.098574

Gonzalez-Guzman, M., Rodriguez, L., Lorenzo-Orts, L., Pons, C., Sarrion-Perdigones, A., Fernandez, M. A., et al. (2014). Tomato PYR/PYL/RCAR abscisic acid receptors show high expression in root, differential sensitivity to the abscisic acid agonist quinabactin, and the capability to enhance plant drought resistance. *J. Exp. Bot.* 65, 4451–4464. doi: 10.1093/jxb/eru219

Hao, Q., Yin, P., Li, W., Wang, L., Yan, C., Lin, Z., et al. (2011). The molecular basis of ABA-independent inhibition of PP2Cs by a subclass of PYL proteins. *Mol. Cell* 42, 662–672. doi: 10.1016/j.molcel.2011.05.011

Hewage, K. A. H., Yang, J. F., Wang, D., Hao, G. F., Yang, G. F., and Zhu, J. K. (2020). Chemical Manipulation of Abscisic Acid Signaling: a New Approach to Abiotic and Biotic Stress Management in Agriculture. *Adv. Sci.* 7:2001265. doi: 10.1002/advs.202001265

Kabsch, W. (2010). Xds. *Acta Crystallogr. D Biol. Crystallogr.* 66, 125–132.

Kai, W., Wang, J., Liang, B., Fu, Y., Zheng, Y., Zhang, W., et al. (2019). PYL9 is involved in the regulation of ABA signaling during tomato fruit ripening. *J. Exp. Bot.* 70, 6305–6319. doi: 10.1093/jxb/erz396

Lin, Z., Li, Y., Wang, Y., Liu, X., Ma, L., Zhang, Z., et al. (2021). Initiation and amplification of SnRK2 activation in abscisic acid signaling. *Nat. Commun.* 12:2456. doi: 10.1038/s41467-021-22812-x

Lin, Z., Li, Y., Zhang, Z., Liu, X., Hsu, C. C., Du, Y., et al. (2020). A RAF-SnRK2 kinase cascade mediates early osmotic stress signaling in higher plants. *Nat. Commun.* 11:613. doi: 10.1038/s41467-020-14477-9

Lozano-Juste, J., Garcia-Maquilon, I., Brea, J., Pina, R., Albert, A., Rodriguez, P. L., et al. (2021). Identification of ABA Receptor Agonists Using a Multiplexed High-Throughput Chemical Screening. *Methods Mol. Biol.* 2213, 99–111. doi: 10.1007/978-1-0716-0954-5_9

Lozano-Juste, J., Garcia-Maquilon, I., Ruiz-Partida, R., and Rodriguez, P. L. (2020). Drug Discovery for Thirsty Crops. *Trends Plant Sci.* 25, 844–846. doi: 10.1016/j.tplants.2020.07.001

Melcher, K., Ng, L. M., Zhou, X. E., Soon, F. F., Xu, Y., Suino-Powell, K. M., et al. (2009). A gate-latch-lock mechanism for hormone signalling by abscisic acid receptors. *Nature* 462, 602–608. doi: 10.1038/nature08613

Mesa-Torres, N., Yunta, C., Fabelo-Rosa, I., Gonzalez-Rubio, J. M., Sanchez-Ruiz, J. M., Salido, E., et al. (2014). The consensus-based approach for gene/enzyme replacement therapies and crystallization strategies: the case of human alanine-glyoxylate aminotransferase. *Biochem. J.* 462, 453–463. doi: 10.1042/BJ20140250

Miao, C., Xiao, L., Hua, K., Zou, C., Zhao, Y., Bressan, R. A., et al. (2018). Mutations in a subfamily of abscisic acid receptor genes promote rice growth and productivity. *Proc. Natl. Acad. Sci. U.S.A.* 115, 6058–6063. doi: 10.1073/pnas.1804774115

Miyazono, K., Miyakawa, T., Sawano, Y., Kubota, K., Kang, H. J., Asano, A., et al. (2009). Structural basis of abscisic acid signalling. *Nature* 462, 609–614. doi: 10.1038/nature08583

Moreno-Alvero, M., Yunta, C., Gonzalez-Guzman, M., Lozano-Juste, J., Benavente, J. L., Arbona, V., et al. (2017). Structure of Ligand-Bound Intermediates of Crop ABA Receptors Highlights PP2C as Necessary ABA Co-receptor. *Mol. Plant* 10, 1250–1253. doi: 10.1016/j.molp.2017.07.004

Mosquna, A., Peterson, F. C., Park, S. Y., Lozano-Juste, J., Volkman, B. F., and Cutler, S. R. (2011). Potent and selective activation of abscisic acid receptors in vivo by mutational stabilization of their agonist-bound conformation. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20838–20843. doi: 10.1073/pnas.1112838108

Nishimura, N., Hitomi, K., Arvai, A. S., Rambo, R. P., Hitomi, C., Cutler, S. R., et al. (2009). Structural mechanism of abscisic acid binding and signaling by dimeric PYR1. *Science* 326, 1373–1379. doi: 10.1126/science.1181829

Okamoto, M., Peterson, F. C., Defries, A., Park, S. Y., Endo, A., Nambara, E., et al. (2013). Activation of dimeric ABA receptors elicits guard cell closure, ABA-regulated gene expression, and drought tolerance. *Proc. Natl. Acad. Sci. U.S.A.* 110, 12132–12137. doi: 10.1073/pnas.1305919110

Pizzio, G. A., Rodriguez, L., Antoni, R., Gonzalez-Guzman, M., Yunta, C., Merilo, E., et al. (2013). The PYL4 A194T mutant uncovers a key role of PYR1-LIKE4/PROTEIN PHOSPHATASE 2CA interaction for abscisic acid signaling and plant drought resistance. *Plant Physiol.* 163, 441–455. doi: 10.1104/pp.113.224162

Rodriguez, P. L., Lozano-Juste, J., and Albert, A. (2019). PYR/PYL/RCAR ABA receptors. *Abscisic. Acid. Plants* 92, 51–82. doi: 10.1016/bs.abr.2019.05.003

Rodriguez-Larrea, D., Perez-Jimenez, R., Sanchez-Romero, I., Delgado-Delgado, A., Fernandez, J. M., and Sanchez-Ruiz, J. M. (2010). Role of conservative mutations in protein multi-property adaptation. *Biochem. J.* 429, 243–249. doi: 10.1042/BJ20100386

Rubio, S., Rodrigues, A., Saez, A., Dizon, M. B., Galle, A., Kim, T. H., et al. (2009). Triple loss of function of protein phosphatases type 2C leads to partial constitutive response to endogenous abscisic acid. *Plant Physiol.* 150, 1345–1355. doi: 10.1104/pp.109.137174

Santiago, J., Rodrigues, A., Saez, A., Rubio, S., Antoni, R., Dupeux, F., et al. (2009). Modulation of drought resistance by the abscisic acid receptor PYL5 through inhibition of clade A PP2Cs. *Plant J.* 60, 575–588. doi: 10.1111/j.1365-313X.2009.03981.x

Sun, L., Wang, Y. P., Chen, P., Ren, J., Ji, K., Li, Q., et al. (2011). Transcriptional regulation of SlPYL, SlPP2C, and SlSnRK2 gene families encoding ABA signal core components during tomato fruit development and drought stress. *J. Exp. Bot.* 62, 5659–5669. doi: 10.1093/jxb/err252

Takahashi, Y., Zhang, J., Hsu, P. K., Ceciliato, P. H. O., Zhang, L., Dubeaux, G., et al. (2020). MAP3Kinase-dependent SnRK2-kinase activation is required for abscisic acid signal transduction and rapid osmotic stress response. *Nat. Commun.* 11:12. doi: 10.1038/s41467-019-13875-y

Tischer, S. V., Wunschel, C., Papacek, M., Kleigrewe, K., Hofmann, T., Christmann, A., et al. (2017). Combinatorial interaction network of abscisic acid receptors and coreceptors from Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U.S.A.* 114, 10280–10285. doi: 10.1073/pnas.1706593114

Vaidya, A. S., and Cutler, S. R. (2022). Chemical Approaches for Improving Plant Water Use. *Methods Mol. Biol.* 2462, 221–230. doi: 10.1007/978-1-0716-2156-1_17

Vaidya, A. S., Helander, J. D. M., Peterson, F. C., Elzinga, D., Dejonghe, W., Kaundal, A., et al. (2019). Dynamic control of plant water use using designed ABA receptor agonists. *Science* 366:eaaw8848. doi: 10.1126/science.aaw8848

Vaidya, A. S., Peterson, F. C., Eckhardt, J., Xing, Z., Park, S. Y., Dejonghe, W., et al. (2021). Click-to-lead design of a picomolar ABA receptor antagonist

with potent activity in vivo. *Proc. Natl. Acad. Sci. U S A.* 118:e2108281118. doi: 10.1073/pnas.2108281118

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., et al. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* 67, 235–242. doi: 10.1107/S0907444910045749

Yang, J. F., Chen, M. X., Zhang, J. H., Hao, G. F., and Yang, G. F. (2020). Genome-wide phylogenetic and structural analysis reveals the molecular evolution of the ABA receptor gene family. *J. Exp. Bot.* 71, 1322–1336. doi: 10.1093/jxb/erz511

Yang, J. F., Hao, G. F., and Yang, G. F. (2021). Genetic Engineering and Chemical Control Related to Abscisic Acid for Improving Plant Drought Tolerance. *J. Agric. Food Chem.* 69, 3563–3565. doi: 10.1021/acs.jafc.1c01039

Yang, Z., Liu, J., Poree, F., Schaeufele, R., Helmke, H., Frackenpohl, J., et al. (2019). Abscisic Acid Receptors and Coreceptors Modulate Plant Water Use Efficiency and Water Productivity. *Plant Physiol.* 180, 1066–1080. doi: 10.1104/pp.18.01238

Yang, Z., Liu, J., Tischer, S. V., Christmann, A., Windisch, W., Schnyder, H., et al. (2016). Leveraging abscisic acid receptors for efficient water use in Arabidopsis. *Proc. Natl. Acad. Sci. U S A.* 113, 6791–6796. doi: 10.1073/pnas.1601954113

Yin, P., Fan, H., Hao, Q., Yuan, X., Wu, D., Pang, Y., et al. (2009). Structural insights into the mechanism of abscisic acid signaling by PYL proteins. *Nat. Struct. Mol. Biol.* 16, 1230–1236. doi: 10.1038/nsmb.1730

Yunta, C., Martinez-Ripoll, M., Zhu, J. K., and Albert, A. (2011). The structure of Arabidopsis thaliana OST1 provides insights into the kinase regulation mechanism in response to osmotic stress. *J. Mol. Biol.* 414, 135–144. doi: 10.1016/j.jmb.2011.09.041

Zhang, X., Jiang, L., Wang, G., Yu, L., Zhang, Q., Xin, Q., et al. (2013). Structural insights into the abscisic acid stereospecificity by the ABA receptors PYR/PYL/RCAR. *PLoS One* 8:e67477. doi: 10.1371/journal.pone.0067477

Zhang, X., and Mou, Z. (2009). Extracellular pyridine nucleotides induce PR gene expression and disease resistance in Arabidopsis. *Plant J.* 57, 302–312. doi: 10.1111/j.1365-313X.2008.03687.x

Zhang, X., Zhang, Q., Xin, Q., Yu, L., Wang, Z., Wu, W., et al. (2012). Complex structures of the abscisic acid receptor PYL3/RCAR13 reveal a unique regulatory mechanism. *Structure* 20, 780–790. doi: 10.1016/j.str.2012.02.019

Zhang, Y., Li, Q., Jiang, L., Kai, W., Liang, B., Wang, J., et al. (2018). Suppressing Type 2C Protein Phosphatases Alters Fruit Ripening and the Stress Response in Tomato. *Plant Cell Physiol.* 59, 142–154. doi: 10.1093/pcp/pcx169

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership