

Comparative Analysis of Sequence-Structure Function Relationship of the SUN-Domain Protein CaSUN1

Poonam Mishra^{1,2†}, Vijay Wardhan^{1†}, Aarti Pandey¹, Subhra Chakraborty¹, Gunjan Garg² and Niranjana Chakraborty^{1*}

¹National Institute of Plant Genome Research, Jawaharlal Nehru University Campus, Aruna Asaf Ali Marg, New Delhi-110067, India

²School of Biotechnology, Gautam Buddha University, Greater NOIDA, Gautam Budh Nagar, Uttar Pradesh-201308, India

†These authors contributed equally to this work.

*Corresponding author: Niranjana Chakraborty, National Institute of Plant Genome Research, Jawaharlal Nehru University Campus, Aruna Asaf Ali Marg, New Delhi-110067, India, Tel: 00-91-11-26735178; E-mail: nchakraborty@nipgr.ac.in

Receiving date: Oct 07, 2017; Acceptance date: Nov 02, 2017; Publication date: Nov 06, 2017

Copyright: © 2017 Mishra P, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Sad1/UNC-84 (SUN)-domain proteins are residents of inner nuclear membrane (INM), and share structural features across species. We previously reported a highly conserved C-terminal SUN-domain family protein, designated CaSUN1, in the stress-responsive proteomic landscape of a grain legume, chickpea. In this study, we identified two other chickpea SUN proteins, CaSUN2 and CaSUN3, and performed a comparative analysis of the sequence-structure-function relationship to better understand the diversification of SUN-domain superfamily proteins. Sequence similarity across the species was investigated using multiple sequence alignment, which showed conserved patterns between CaSUN1 and the homologs. Phylogenetic analysis showed that plant SUN-domain proteins are clustered in a unique and distinct group. Using ab-initio approach, a 3D protein structure was generated and further validated using various tools including the Ramachandran plot. The results displayed 90.1% of the ϕ and ψ residues angles in the most favoured regions, suggesting a high-quality structural model for CaSUN1. Model deviation and fluctuation analysis were performed using molecular dynamics (MD) simulation of CaSUN1. The secondary structure analysis of CaSUN revealed a similarity between the structural components shared among them. CaSUN1 revealed two functional domains viz., SUN and muskellin, and the presence of kelch-repeat domain pointed out its putative role in oligomerization, while its binding affinity with different ligands indicates diverse functions. These results would not only give deeper insights into the structure-function relationships within the SUN-superfamily proteins, but also their putative physiological roles.

Keywords: 3-dimensional modelling; Grain legume; MD simulation; Multiple sequence alignment; Phylogenetic analysis; Ramachandran plot; SUN-domain

Introduction

Plants, due to their sessile nature, are exposed to many environmental stresses which continuously influence growth and development and reduce crop productivity. Water-deficit or dehydration is the most widespread manifestation of environmental stress as it is associated with every abiotic stress by one way or another, and is detrimental to plant health in itself as well as exacerbates the effects of other stresses. We previously explored the dehydration-responsive membrane-associated proteome of chickpea and identified a wide group of proteins, including a non-canonical SUN-domain protein, designated CaSUN1 (*Cicer arietinum* SUN1) [1]. The SUN-domain superfamily proteins have a conserved C-terminal motif of few hundred amino acids (~200 aa), representing the SUN-domain (Sad1 and UNC-84 domain). These proteins show homology with the *Caenorhabditis elegans* protein UNC-84, an important nuclear envelope protein implicated in nuclear anchoring and/or positioning during development. The other homologous protein, Sad1 in *Schizosaccharomyces pombe*, localizes to the spindle pole body but is also known to localize at the nuclear envelope when overexpressed. It plays an important role in DNA duplication by anchoring centrosome and spindle body to the nuclear envelope, and participates in transferring mechanical force, generated in cytoplasm, to inner nuclear

membrane. Characteristics feature of this superfamily is the presence of SUN-domain as well as, at least, one transmembrane domain. The occurrence of two subfamilies of SUN-domain proteins in flowering plants and moss suggests its early origination during evolution. The discovery of this gene superfamily provides a valuable insight into the nuclear positioning events during various developmental stages in the plant life cycle. Further investigation on molecular mechanisms would be significant for deciphering the biological role of the plant SUN proteins.

The molecular aspects associated with vital cellular processes can be elaborated using evolutionary studies of specific metabolic pathways across organisms. A broad range of functions is linked with the interaction of various nucleoplasmic and nuclear membrane proteins with lamins, and lamin-associated protein complexes. Several of these nuclear membrane proteins contain distinct protein domain structures leading to their categorisation as LEM-, SUN- or KASH-domain proteins. Nuclear envelope localization of some SUN-KASH-domain proteins has been attributed to their interaction with lamins. These proteins also serve as specialized nuclear envelope receptors, which interconnect the cytoskeleton and nucleoskeleton. The direct interaction between SUN-domain and KASH-domain is known to influence the cellular functions of both of these proteins [2,3]. SUN-domain proteins have been found to be increasing in number over the course of evolution as evident by the presence of single SUN-domain gene in *S. pombe* genome, two genes each in *Caenorhabditis elegans*

and *Drosophila melanogaster*, whereas four or more SUN-domain genes in mammals.

The members of SUN superfamily, besides the SUN-domain, share various other structural features to different extents. According to their function, SUN-domain proteins possess variable number of transmembrane domains (TMDs) with a minimum of one TMD across all members. The multiple TMDs are proposed to transmit mechanical force by anchoring mechanical-load-bearing structures, as in the case of human SUN1, which spans the membrane three times [4]. Apart from them, SUN family proteins contain additional conserved hydrophobic region/s, which may or may not span the membrane [5,6]. Interestingly, TMD lacking SUN-domain proteins have been reported in *D. melanogaster*. Another important property of SUN-domain proteins is the presence of, at least, one coiled-coil domain, which might have a role in mediating homodimerization [4]. The possibility of dimerization could be an important strategy for SUN-domain proteins as each SUN-domain dimer could in turn directly anchor two KASH-domain partners giving rise to higher-order filamentous structures bridging the nuclear envelope. The interaction between the KASH-domain with SUN-domain proteins not just provides a mechanism to connect several nucleoskeletal and cytoskeletal structures, but is also implicated in nuclear positioning apart from other non-mechanical roles. For example, SUN-domain and KASH-domain proteins have been shown to be involved in chromosome organization and dynamics during meiosis and abiotic stress responses, among others.

In the present study, we have analyzed the amino acid sequence of CaSUN1 with respect to other plant SUN proteins. The extensive bioinformatics analysis uncovered new knowledge about the structural features of CaSUN proteins. This investigation highlights the utility of sequence-based evaluation in providing a structural basis for the protein function analysis of plant SUN proteins. We carried out a 3D simulation to obtain biologically relevant structural models of CaSUN1, besides establishing its evolutionary history based on phylogenetic analysis. The *in silico* 3D structure prediction of CaSUN1 is particularly helpful in view of the assumption that structure is more conserved than sequence between homologous proteins. Model deviation (MD) structure of CaSUN1 describes not only its shape, but also its characteristics, which affect the hydrophobicity and hence its function. These approaches provide primary assessment for the functional role of previously uncharacterized proteins.

Materials and Methods

Data set

The amino acid sequence of CaSUN1 was retrieved from NCBI Database (Accession number: XP_004516032.1). This sequence was used to assort similar sequences having known 3D structures against Protein Data Bank (PDB) [7] by NCBI's BLASTp program.

Primary structure prediction

Primary structure prediction was done using ExPASy [8], and ProtParam server which analyses physiochemical characters of CaSUN1 such as theoretical isoelectric point (pI), molecular mass, molecular formula, instability index, total number of positive and negative residues, aliphatic index and grand average of hydropathicity (GRAVY).

Secondary structure prediction

Secondary structure prediction was done using PSIPRED [9] server, which analyses the total number of α -helix, β -sheet and turns.

Multiple sequence alignment

The amino acid sequence of CaSUN1 and their homologs were subjected to multiple sequence alignment (MSA) to recognize the conserved residues and patterns using MEGA-4 program [10]. The conserved residues were used for further analysis. The gap penalties alignment parameters was -11 to -1, end gap penalties was -5 to -1, and e-value was 0.003, word size was 4 and maximum cluster distance was taken as 0.8. The motifs of the SUN proteins were identified using MEME online tools (<http://meme.nbcr.net/meme/>). The parameters taken were: the maximum length of the conserved motif was 50, the minimum length of the conserved motif was 6, and the largest number of discovered and conserved motifs was 15.

Phylogenetic analysis

MSA of amino acids was performed with MEGA-4 program [10]. A phylogenetic tree was constructed by neighbor-joining method using the default parameters.

Subcellular localization and promoter analysis

Subcellular localization was predicted using the WoLF PSORT (http://www.genscript.com/psort/wolf_psort.html), BaCellLo (<http://gpcr.biocomp.unibo.it/bacello>), ESLPred (<http://www.imtech.res.in/raghava/eslpred/>), YLoc (<http://abi.inf.uni-tuebingen.de/Services/YLoc/webloc.cgi>), PSORT II (<http://psort.hgc.jp/form2.html>), LocTree3 (<https://roslab.org/services/loctree2/>), Plant-mPLOC (<http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/>) and CELLO (<http://cello.life.nctu.edu.tw/>). The NCBI was used to retrieve a 1500 bp DNA sequence from 5'-upstream region of each *CaSUN* gene (<http://www.ncbi.nlm.nih.gov/>) and were subjected to the plantCARE database (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) for cis-element scan.

Prediction of miRNA targets

The potential miRNA targets were predicted with psRNATarget program (<http://bioinfo3.noble.org/psRNATarget/>) using default parameters. The sequences of newly identified miRNAs and significantly different known miRNAs were used as custom sequences. The redundant sequences were removed after identifying potential target mRNA sequences for functional analysis.

3D structure prediction

3D structure prediction was done using ab initio approach, and model was generated using Modeller script. A representative PDB structure library was scanned against PDB using NCBI's BLASTp program to search for template structure and alignment. I-TASSER programme [11] was also used for template search and 3D structure analysis. The templates were used to identify the existence of similar regions from different models. The 3D structure of all the identified homologs were downloaded from PDB database and used for structure prediction.

homology and hence their homologs in plants have not been identified by sequence analysis thus far. Identification of SUN-interacting partners by molecular interaction screens may be required for elucidating their function in plants.

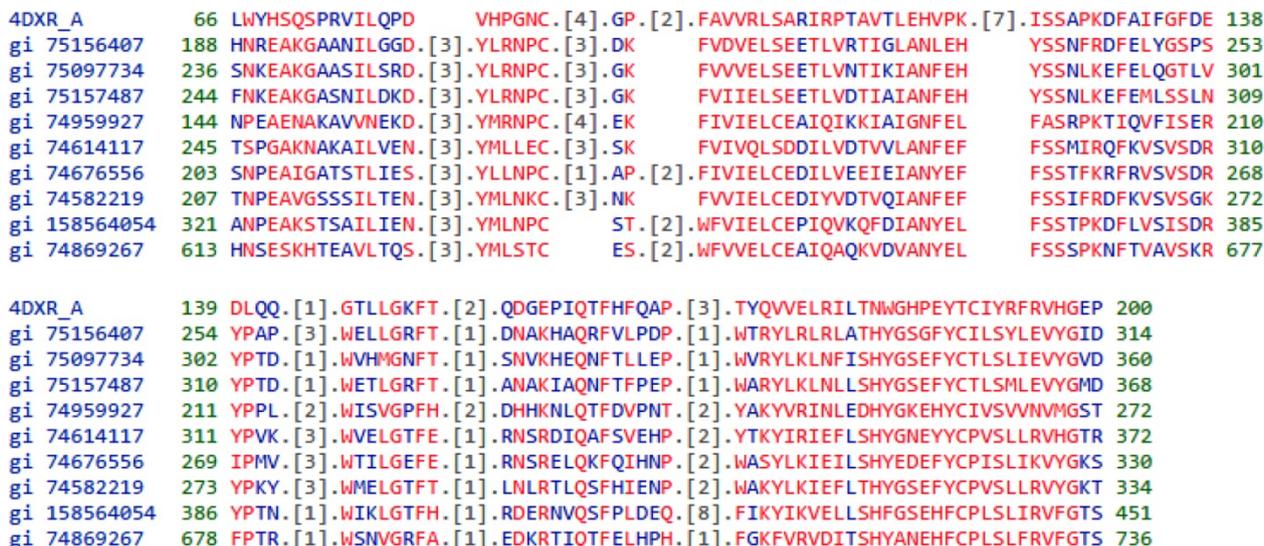


Figure 2: The multiple sequence alignments of SUN-domain proteins show the conservation of residues across species. Blue colour represents the conserved residues and red colour indicates the identical residues with their respective positions.

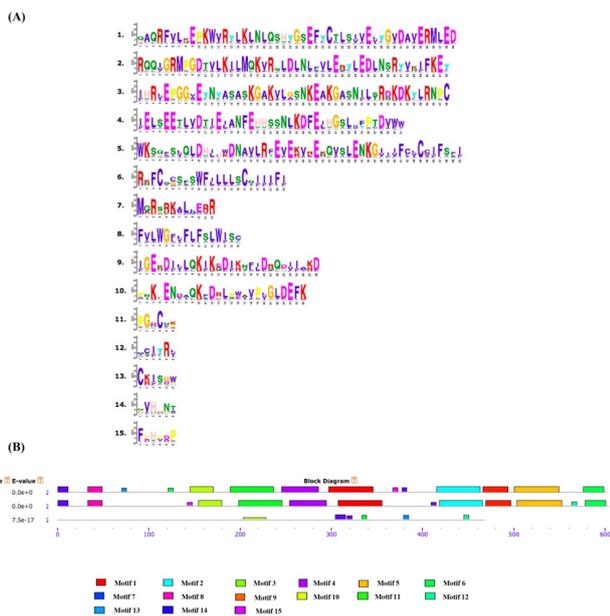


Figure 3: Schematic representation of the conserved motifs in chickpea SUN proteins (A). Each colored box represents a motif in the protein (B), with the motif name indicated for the box on the right. The length of the protein and motif can be estimated using the scale at the bottom.

Phylogenetic analysis

Phylogenetic and evolutionary relationships of SUN-domain proteins from different organisms was investigated using their full length protein sequences, which lead to the construction of an unrooted phylogram. The phylogenetic analysis revealed major clusters as shown in Figure 4. The SUN proteins from plants, yeasts and animals formed distinct clusters. Among the plants, monocots and dicots formed separate clade. The CaSUN1 showed high homology to Medicago and other legumes and thus grouped together. Cluster I largely comprised of legume species, while Cluster II contained other plant species. Likewise, the yeasts were out-grouped from Cluster II and formed Cluster III. Cluster IV was composed of SUN-domain proteins from animal species and were grouped together. This shows a conservation of the SUN superfamily proteins among the close species, while diversification during the course of evolution.

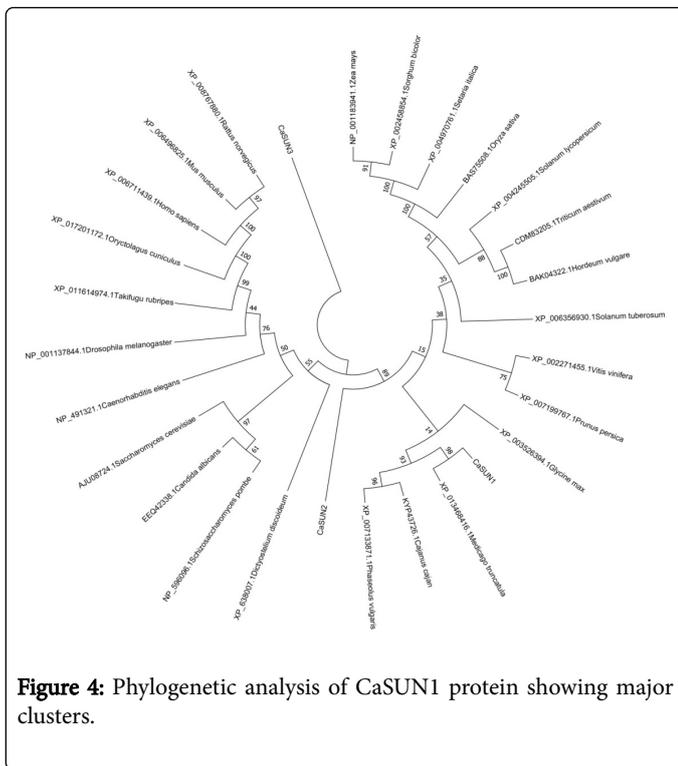


Figure 4: Phylogenetic analysis of CaSUN1 protein showing major clusters.

Analysis of upstream and downstream regulatory elements associated with chickpea SUN proteins

To analyze the regulation of CaSUN genes under stress conditions, we evaluated 1500-bp sequences upstream of the transcriptional start site (Supplementary Table S2). The identified putative cis-acting

regulatory elements (CAREs) revealed enhancer, essential, hormone-responsive, stress-responsive, and other elements. Among the essential CAREs, TATA and CAAT boxes were detected. The CAREs associated with environmental stress (MBS, HSE, TGA, TCA and TC-rich repeats) and hormone response (CGTCA-motif, TGACG element) were identified. While HSE and MBS involved in water-deficit were detected in promoters of all the three chickpea SUN proteins; TCA-element involved in SA response and MeJA responsive CGTCA-motif were found exclusively in CaSUN1. Similarly, the environmental stress responsive 5UTR PY-rich stretch and TC-rich repeats were found in CaSUN2, while TGA-elements were predicted in CaSUN3. For the downstream regulation of CaSUN proteins we analysed their sequenes for their target miRNA. The miRNAs are small, endogenous RNAs, which dictate gene expression during plant development and stress responses at post-transcriptional level. Under stress conditions many genes have been reported to be regulated post-transcriptionally through several miRNA families [19]. Among CaSUN genes, CaSUN1 showed the target sequences for miR2873a, miR7741-5p.1 and miR8743b, while CaSUN2 revealed target sequences for miR3949 and miR5674a (Supplementary Table S3). The expression of miR5674 was found to be upregulated in *E. culmorum* during environmental stress [19].

3D structure prediction and homology modeling

Non-availability of crystal structure has been a major hurdle in prediction of functional importance of the conserved motifs. Hence, for deciphering the functional significance of CaSUN proteins, known crystal structure were taken as a template for mapping the conserved residues. The complete 3D structure of CaSUN1 was predicted using appropriate template via ab initio based modeling. The best aligned template (PDB ID: 5ED8 Chain-A) was obtained by searching against PDB database, where sequence identities was 20.42%, and query coverage was 196 to 338 amino acid residues.

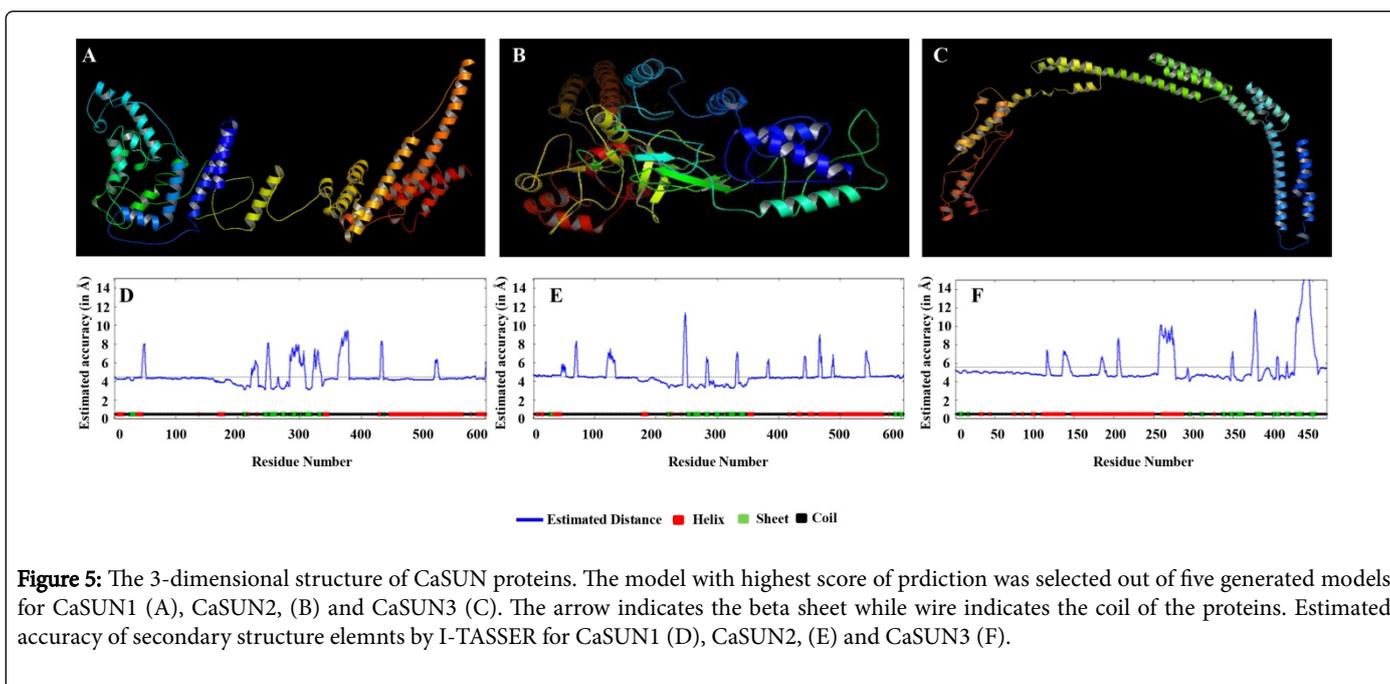


Figure 5: The 3-dimensional structure of CaSUN proteins. The model with highest score of prdiction was selected out of five generated models for CaSUN1 (A), CaSUN2, (B) and CaSUN3 (C). The arrow indicates the beta sheet while wire indicates the coil of the proteins. Estimated accuracy of secondary structure elemnts by I-TASSER for CaSUN1 (D), CaSUN2, (E) and CaSUN3 (F).

The templates were selected as the targets and for the analysis of 3D structure. A high level of sequence identity/similarity and accurate alignment between the targets and template led to the prediction of

five models generated with the target sequence. The highly predicted models were selected using QMEAN score value (-5.86) for the best predicted model. Additionally, we used I-TASSER programme for

homology modelling. To generate the 3D structure of CaSUN1, the 602 amino acid long sequence was submitted to LOMETS [20,21]. The best structural template predicted for CaSUN1 was a human nuclear pore protein (PDB ID 5a9qA). The best 3D structure predicted was aligned to 5a9qA (human nuclear pore complex) in TM-align server [22] and checked for accuracy. The resultant model of CaSUN1 was examined using Chimera 1.2 [23], which showed 11 α -helices, 9 β -sheets, and 19 coils (Figure 5A). Similarly, the CaSUN2 and CaSUN3 Protein were best aligned with 4dxtA and 3unpA, respectively, (Figure 5B-C). Although the sequence identity was significantly lower, the alignment of the template-structure showed a significant match. In threading, the percentage sequence identity for the aligned region of the templates with the query sequence remained between 0.07 and 0.10, while on the whole it was 0.19. Coverage of threading alignment ranged between 0.96 and 0.98. The normalized Z-scores of the threading alignments ranged between 2.06 and 0.96. These parameters for CaSUN2 and CaSUN3 were more consistent suggesting a significant threading alignment with their respective templates (Table 1). The accuracy of the predicted models for CaSUN proteins by I-TASSER was estimated using C-score, which gave a value of 0.49, 0.64 and 1.16 for CaSUN1, CaSUN2 and CaSUN3, respectively. The values of other parameters (number of decoys and cluster density) used for the structure validation were also in the reliable range. TM-score for CaSUN1 was 0.65, while estimated RMSD was 8.9. A TM-score >0.5 specifies the correct topology of the predicted model, while a TM-score <0.17 indicates only random similarity. The TM-score of 0.63 and 0.57 along with an RMSD value of 9.2 and 9.8 for the predicted model of CaSUN2 and CaSUN3, respectively, indicates its highly probable 3D structure with a correct topology (Table 2 and Figure 5D-F).

Positioning of donor and acceptor molecules

The program I-TASSER offers information about the possible ligand binding sites. For CaSUN1, the top prediction for a structural template for binding site was 1ocoB, a bovine heart cytochrome C oxidase in carbon monoxide-bound state at val35 and lys36, which exhibited a C-score of 0.07. The alignment between 1ocoB and the predicted CaSUN1 model also identified some key binding residues. Thr 84, Glu258 were found to be involved in Zn⁺ binding, while Gln318, Phe321, and Gly322 displayed active role in binding urea. Similarly CaSUN2 and CaSUN3 also predicted their binding with various ligands such as FWD, AKG, peptides and nucleotides, among others, indicating their diverse cellular functions. The binding of CaSUN1 with these ligands suggests other possible roles of SUN-domain proteins than merely anchorage in nuclear pore complex.

3D structure validation and annotation

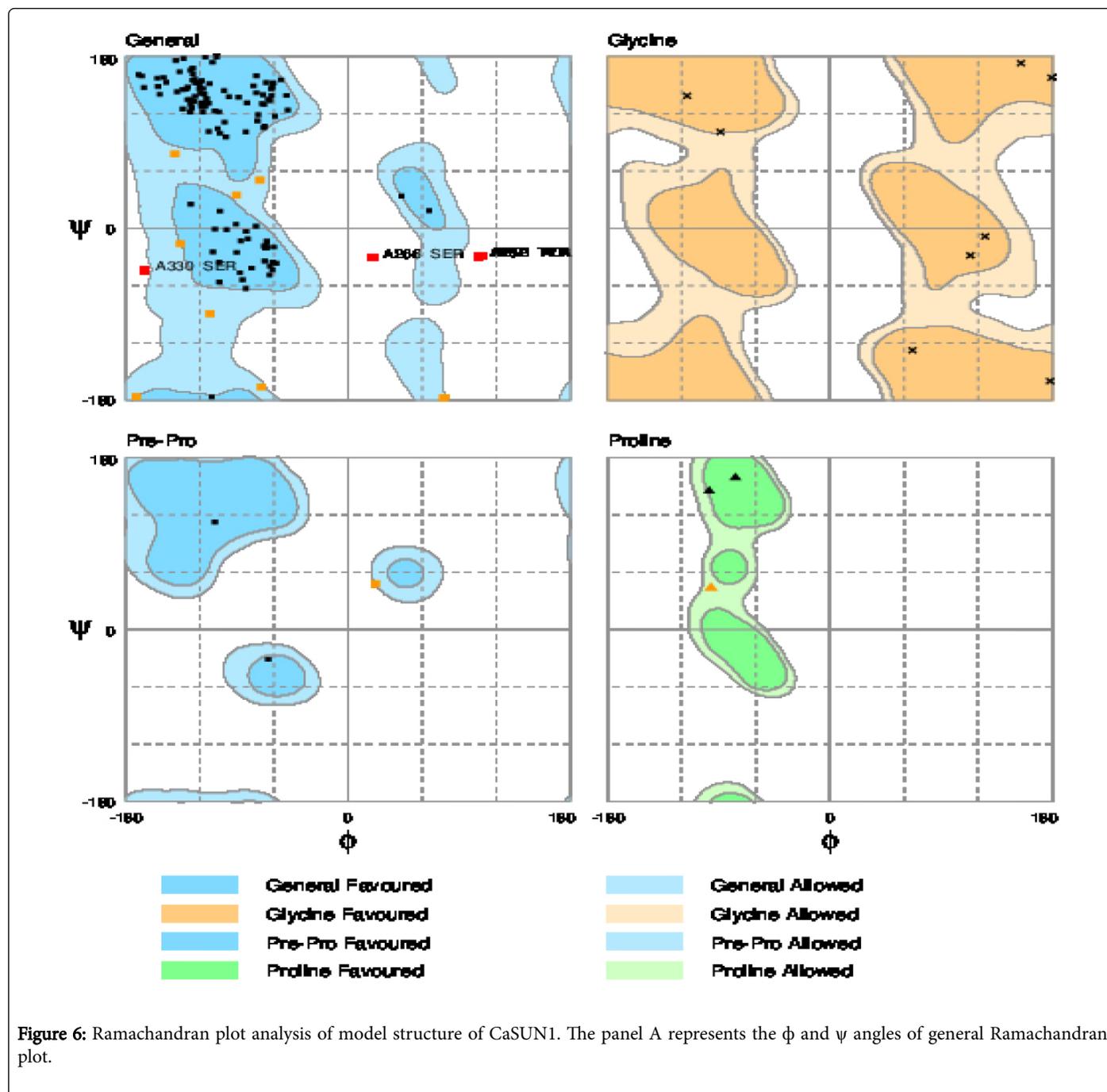
Among the three chickpea SUN proteins CaSUN1 has been found to be involved in abiotic stress response in plants [1] and therefore to investigate the structure-function relationship of this protein, we further validated its predicted 3D structure. The 3D model so acquired was visualized in PyMol program (<http://www.pymol.org>) and evaluated on the basis of Ramachandran plot obtained by PSVS analysis. The Ramachandran plot of a particular protein serves as an important indicator of the quality of 3D structures. Further, the analysis of the distribution of torsion angles (ϕ and ψ angles) in a protein structure provides an important local structural parameter that control protein folding. The Ramchandran plot for CaSUN1 showed that the ϕ and ψ angles of 90.1% of the residues fell in the most favoured regions followed by 7.1% residues in the allowed regions. A comparatively miniscule quantity of residues (~2.8%) was found in the disallowed regions (Figure 6). The occurrence of more than 90% residues in the most favoured region suggests that the structural model produced is a presumably high quality model for CaSUN1.

Protein	PDB Template	Iden1 ^a	Iden2 ^b	Covc	Norm. score ^d	Z-
CaSUN1	5a9qA	0.07-0.10	0.19	0.96-0.98	0.96-2.06	
CaSUN2	4dxtA	0.20-0.21	0.06	0.27-0.28	1.35-3.82	
CaSUN3	3unpA	0.32-0.33	0.14	0.40-0.41	1.14-2.50	

Table 1: Predicted templates for threading by I TASSER. Ident1 is the percentage sequence identity of the templates in the threading aligned region with the query sequence. Ident2 is the percentage sequence identity of the whole template chains with query sequence. Cov represents the coverage of the threading alignment and is equal to the number of aligned residues divided by the length of query protein. Norm. Z-score is the normalized Z-score of the threading alignments.

Protein	C-score	Estimated TM-score	Estimated RMSD
CaSUN1	0.49	0.65 ± 0.13	8.9 ± 4.6Å
CaSUN2	0.64	0.63 ± 0.13	9.2 ± 4.6Å
CaSUN3	1.16	0.57 ± 0.15	9.8 ± 4.6Å

Table 2: Parameters for Predicted 3D model through I TASSER.



MD simulation

Root mean square deviation (RMSD) and root mean square fluctuation (RMSF) analyses were used to identify simulation equilibration as well as fluctuations of CaSUN1 from starting to completed MD simulation.

RMSD was applied to analyse the average variation atoms for a specific frame which is calculated from the reference frames in a trajectory. The value of RMSD for x frame was calculated from the following equation:

$$RMSD_x = 1/N \sqrt{\left(\sum_{i=1}^N R'_i(Tx) \right) - (R_j(Tref))^2}$$

Where N in the above equation represents total number of atoms, R' is position of the selected atoms where frame x is recorded at time Tx and Tref is the reference time. Same procedure was reiterated for all frames in the simulation trajectory. The information regarding structural conformation throughout the simulation can be obtained by monitoring the RMSD of the protein. A fluctuation around some thermal average structures has been generally observed towards the end of simulation. The equilibration in the simulation is indicated by RMSD value. A variation of the order of 1-3 Å are acceptable for small

and globular proteins, while changes much larger than that indicate a large conformational change in the protein structure during the simulation [24]. CaSUN1 frames were aligned with respect to reference frame of backbone, and calculated RMSD was done by atom selection. MD simulation value was stabilized with some fluctuation at later stage for whole simulation throughout the 5 ns trajectory (Figure 7). The

RMSD value was found to be below 3 Å, which indicated a stable conformation of CaSUN1 backbone with the template. Similarly, less than 3.5 Å values of RMSD for side chains and heavy atoms across time suggest a valid structure for CaSUN1 as predicted against the known template.

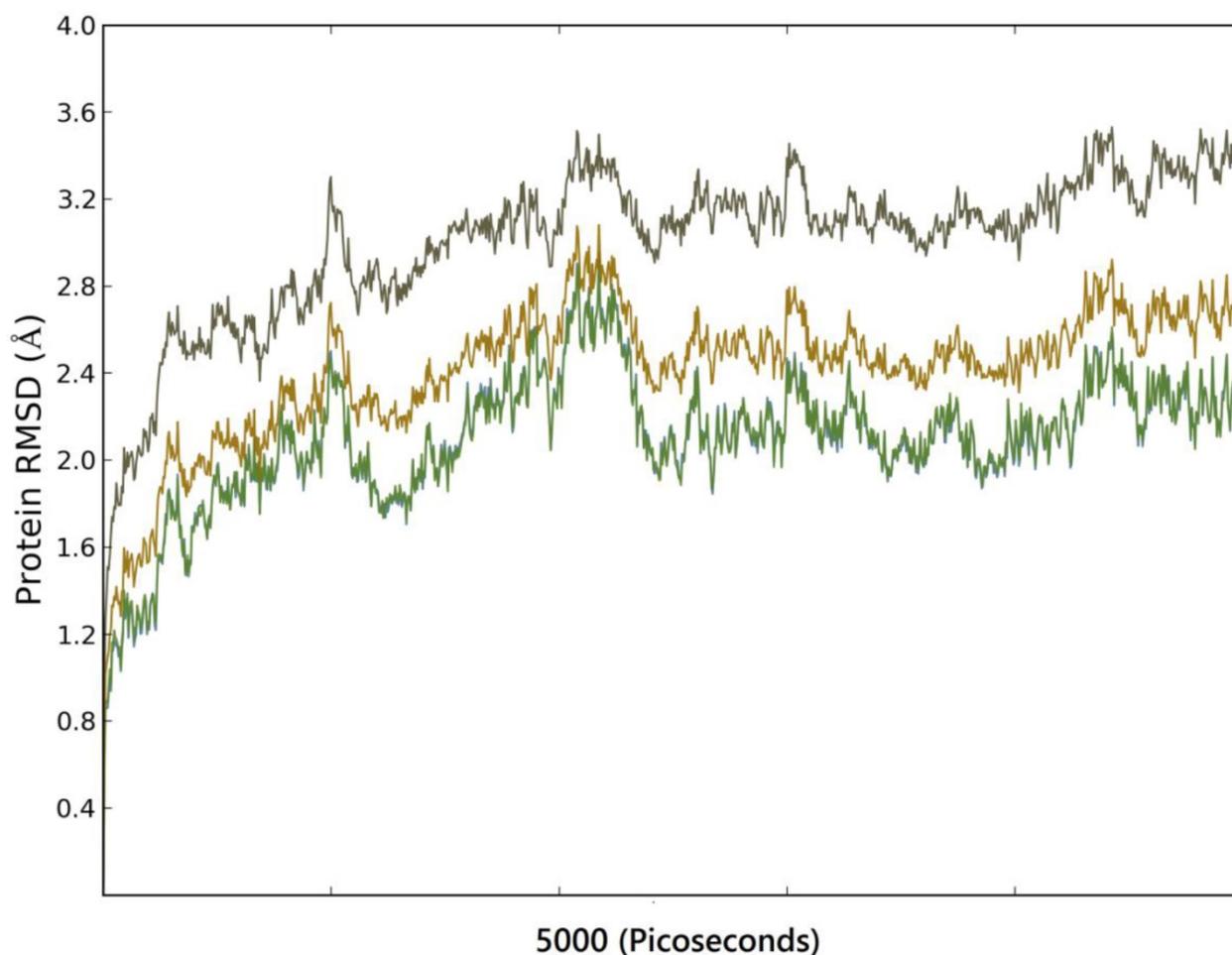


Figure 7: RMSD vs. time plot of MD simulation interaction graph for CaSUN1. The plot represents the RMSD evolution of CaSUN1. Green peaks represent the backbone, black peaks represent the side chain and brown peaks represent the heavy atoms of CaSUN1.

RMSF helped in investigating local changes along the CaSUN1 protein chain. RMSF for i:

$$RMSFi = \frac{1}{T} \sqrt{\left\langle \sum_{t=1}^T (R_i(t) - R_j(T_{ref}))^2 \right\rangle}$$

Where T is trajectory time calculated over RMS fluctuation, Ri is the position of residue i, R' is the position of atoms in residue i, Tref is the reference time and the angle brackets represents the average of the

square distance. RMSF was performed to observe the flexibility of different segments of CaSUN1. We monitored RMSF values for each residue of CaSUN1 in time averaged position. The RMSF plot showed a high peak at the N-terminal region, while comparatively lower peak in other residues (Figure 8). The lowest peaks in the C-terminal region represent least fluctuated residues of the CaSUN1 protein during the MD simulation. The results indicate more rigid secondary structure elements than the unstructured part of the protein and a less fluctuated structure on loop or turn regions.

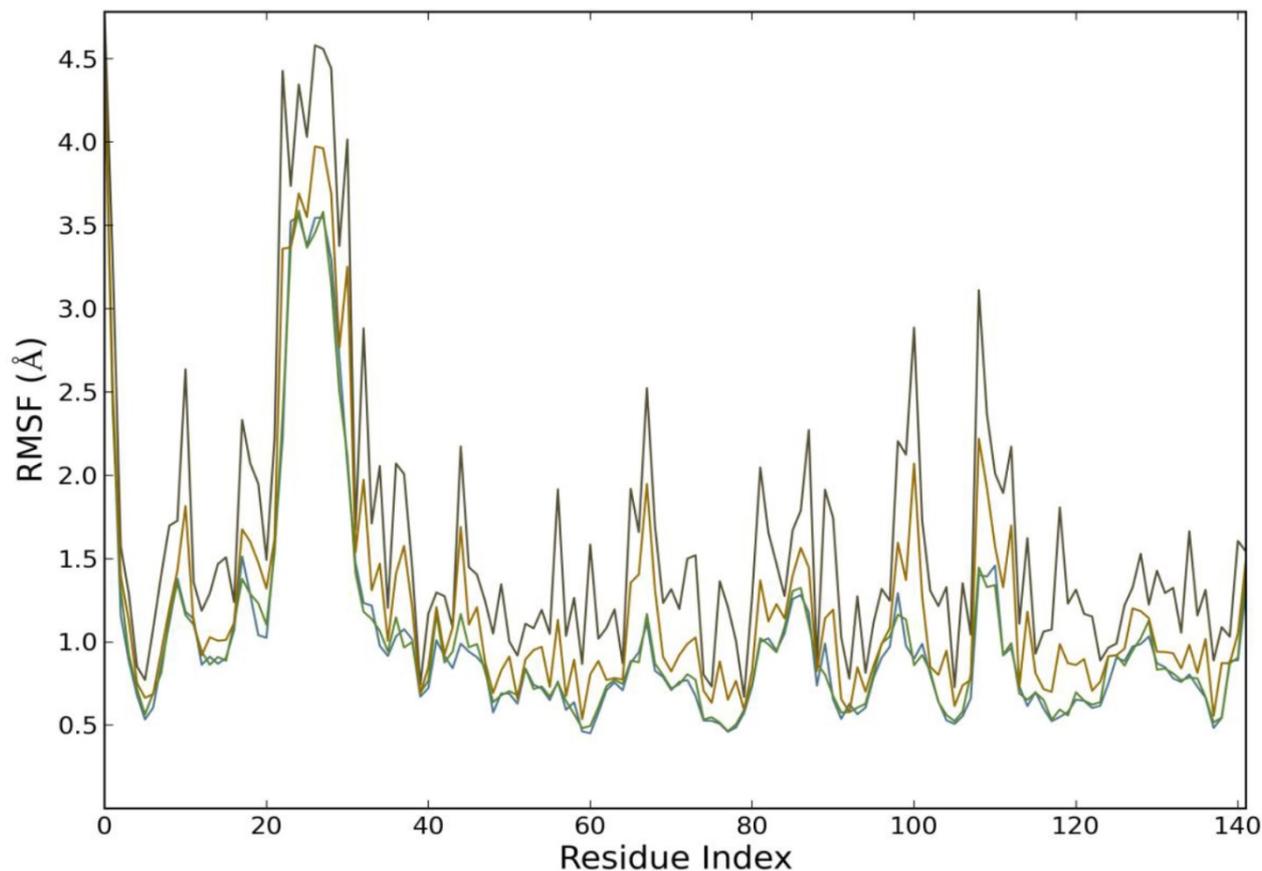
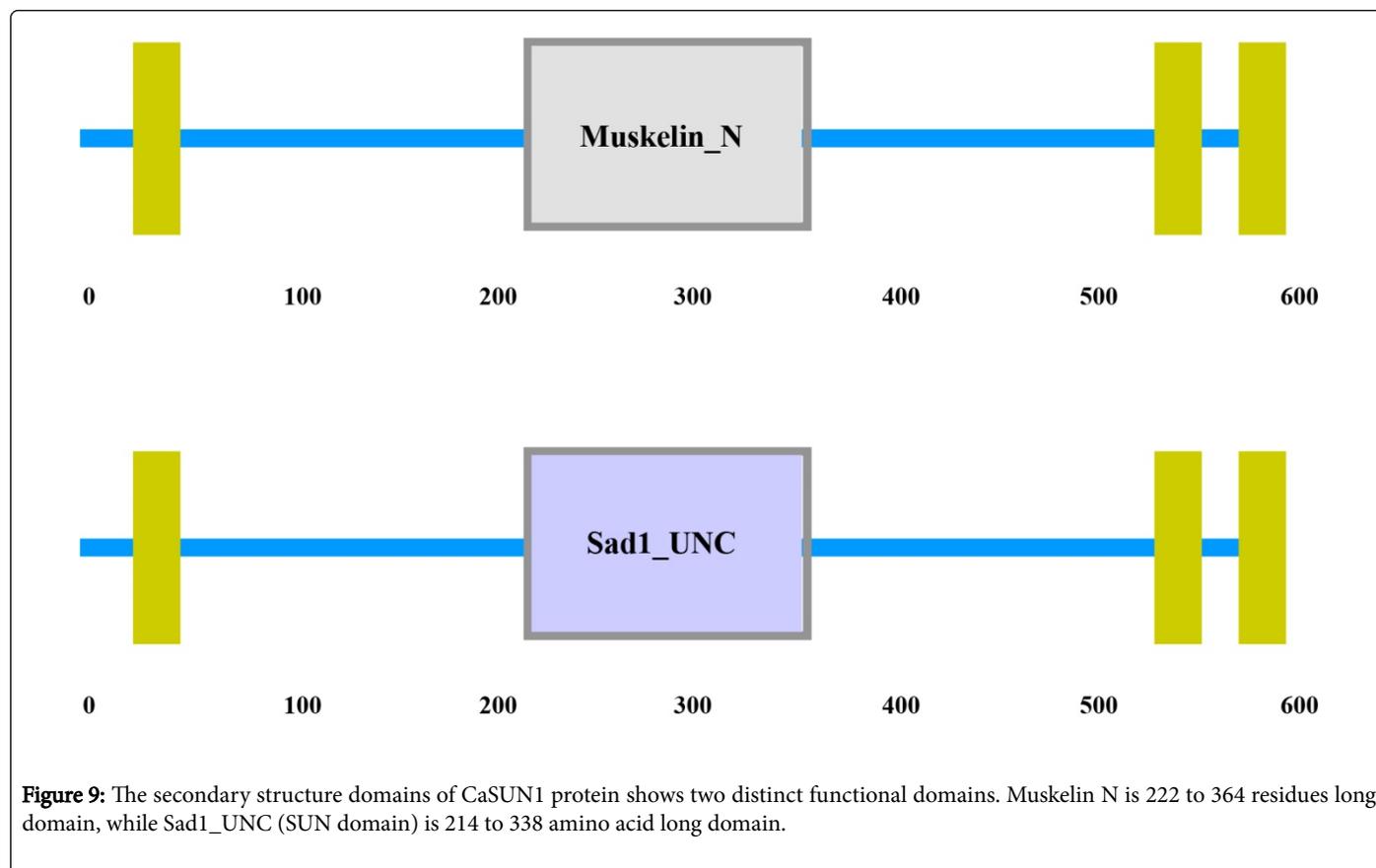


Figure 8: RMSF vs. residue plot of MD simulation graph for CaSUN1. Blue peaks represent the alpha-carbon, green peaks represent the backbone, black peaks represent the side chain and brown peaks represent the heavy atom of CaSUN1.

Sequence-structure-function relationship

Analysis of conserved patterns and their structural components, as observed in the MSA profile of homologous sequences, established the sequence-structure-function relationship of CaSUN1. The structural roles of identified conserved patterns are highlighted with the descriptions of their domain families (Figure 9). Out of 96 conserved patterns for the domain family used [25], only two conserved patterns were observed for CaSUN1. These patterns found as the member of muskelin_N and Sad1_UNC domain family were predicted to be involved in the function of CaSUN1. The muskelin_N is 222 to 364 residues long domain, while Sad1_UNC is 214 to 338 amino acid long domain. Muskelin was initially identified in vertebrates as a novel intracellular multidomain protein mediating cell spreading responses to the matrix adhesion molecule, thrombospondin-1 [26]. This protein is a kelch-repeat protein. The domain is largely involved in the regulation of ubiquitin-mediated protein degradation and has a role in

actin binding [27,28]. Many kelch-repeat proteins have been shown to have the ability to self-associate into dimers or oligomers. Interestingly, muskelin has not been found in plants or the yeast [29], but in mammals it was found to be associated with the RanBP9 complex [30]. It is a component of a putative E3 ligase complex as well as has a role in cell adhesion and regulation of cytoskeleton dynamics. Previous studies have shown that the two predicted coiled coil motifs in the luminal region near the N-terminal end of the SUN domain are responsible for the formation of homo- or hetero-oligomers in SUN1 and SUN2 [31]. Further analyses of the potential coiled coil motifs within the luminal domain revealed a high probability of dimerization of the first coiled coil, while a modest chance of trimerization for the second one [31]. It is likely that the dimerization and trimerization at distinct regions of the SUN-domain protein luminal region are guided by SUN-domain and the two coiled coil motifs, respectively which would be responsible for its specific functions.



Conclusion

The function of a protein is significantly related to its structure, which in turn is defined by the presence of various functional domains and amino acid residues. The CaSUN1, a SUN-domain protein showed conserved C-terminal SUN-domain when compared with other SUN superfamily proteins from across species. The CaSUN1 and its homologs in chickpea CaSUN2 and CaSUN3 revealed striking evolutionary conservation as well as diversification among plants and other species. A highly predicted 3D structure of CaSUN1 with the available template was verified with Ramachandran plot analysis and MD simulations. CaSUN1 protein sequence suggested a secondary structure with a binding affinity for some ligands, which indicates the functions of this protein other than its role in nuclear architecture. Presence of a domain related to muskelin protein, a kelch-repeat family protein, pointed out its putative role in oligomerisation and hence diverse functions. This work provides an insight into the functional aspects of a plant SUN protein and a myriad of other such proteins with an enhanced understanding of their function associated with their various structural features.

Availability of data and materials

All the analytic programs and bioinformatic databases are freely available with three web browsers: Mozilla Firefox, Internet Explorer, and Safari, and four operating systems: Windows XP, Windows Vista, Linux (Red Hat), and Mac OS. No additional software installation is needed for browsing the databases.

Competing Interests

The authors declare that they have no competing interests.

Funding

This work was supported by SERB grant [EMR/2015/001870 from the Department of Science and Technology (DST)], Govt. of India. The authors also thank DST for providing pre-doctoral fellowship [SR/WOS-A/LS-98/2016 (G)] to P.M. and the Council of Scientific & Industrial Research (CSIR), Govt. of India for providing post-doctoral fellowship [38(1385)/13/EMR-II] to V.W. The authors thank the National Institute of Plant Genome Research, New Delhi for providing post-doctoral fellowship to A.P.

Authors' Contributions

P.M., S.C. and N.C. conceived the project. P.M. and V.W. designed and performed the study. P.M. and V.W. carried out the data analysis. P.M., V.W., A.P., G.G. and N.C. discussed the study and wrote the article. All authors read and approved the final manuscript.

Acknowledgements

We are thankful to Jasbeer Singh for illustrations and graphical representation in the manuscript.

References

1. Jaiswal DK, Mishra P, Subba P, Rathi D, Chakraborty S, et al. (2014) Membrane-associated proteomics of chickpea identifies Sad1/

- UNC-84 protein (CaSUN1), a novel component of dehydration signaling. *Sci Rep* 4: 4177.
2. McGee MD, Rillo R, Anderson AS, Starr DA (2006) UNC-83 Is a KASH protein required for nuclear migration and is recruited to the outer nuclear membrane by a physical interaction with the SUN protein UNC-84. *Mol Biol Cell* 17: 1790-1801.
 3. Padmakumar VC, Libotte T, Lu W, Zaim H, Abraham S, et al. (2005) The inner nuclear membrane protein Sun1 mediates the anchorage of Nesprin-2 to the nuclear envelope. *J Cell Sci* 118: 3419-3430.
 4. Crisp M, Liu Q, Roux K, Rattner JB, Shanahan C, et al. (2006) Coupling of the nucleus and cytoplasm: role of the LINC complex. *J Cell Biol* 172: 41-53.
 5. Hodzic DM, Yeater DB, Bengtsson L, Otto H, Stahl PD (2004) Sun2 is a novel mammalian inner nuclear membrane protein. *J Biol* 279: 25805-25812.
 6. Liu Q, Pante N, Misteli T, Elsagga M, Crisp M, et al. (2007) Functional association of Sun1 with nuclear pore complexes. *J Cell Biol* 178: 785-798.
 7. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28: 235-242.
 8. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, et al. (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31: 3784-3788.
 9. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics* 16: 404-405.
 10. Tamura K, Dudley J, Nei M, Kumar S (2007) Mega4 molecular evolutionary genetics analysis (MEGA) software, version 4.0. *Mol Biol Evol* 24: 1596-1599.
 11. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction *Nat Protoc* 5: 725-738.
 12. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Molecular Graphics* 14: 33-38.
 13. Bhattacharya A, Tejero R, Montelione GT (2007) Evaluating protein structures determined by structural genomics consortia. *Proteins* 66: 778-795.
 14. Luthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. *Nature* 356: 83-85.
 15. Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins* 17: 355-362.
 16. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26: 283-291.
 17. Sonnhammer EL, Eddy SR, Birney E, Bateman A, Durbin R (1998) Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Res* 26: 320-322.
 18. Starr DA, Fischer JA (2005) KASH'n Karry: The KASH domain family of cargospecific cytoskeletal adaptor proteins. *Bioessays* 27: 1136-1146.
 19. Yu CS, Chen YC, Lu CH, Hwang JK (2006) Prediction of protein subcellular localization. *Proteins* 64: 643-651.
 20. Inal B, Turktas M, Eren H, Ilham E, Okay S, et al. (2014) Genome-wide fungal stress responsive miRNA expression in wheat. *Planta* 240: 1287-1298.
 21. Wu S, Zhang Y (2007) LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res* 35: 3375-3382.
 22. Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 33: 2302-2309.
 23. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera - a visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605-1612.
 24. Frenkel D, Smit B (1996) Understanding Molecular Simulations: From Algorithms to Applications, Academic Press, San Diego.
 25. Avila CL, Rapisarda VA, Farias RN, De Las Rivas J, Chehín R (2007) Linear array of conserved sequence motifs to discriminate protein subfamilies: study on pyridine nucleotide-disulfide reductases. *BMC bioinformatics* 8: 96.
 26. Adams JC, Seed B, Lawler J (1998) Muskelin, a novel intracellular mediator of cell adhesive and cytoskeletal responses to thrombospondin1. *EMBO J* 17: 4964-4974.
 27. Adams J, Kelso R, Cooley L (2000) The kelch repeat superfamily of proteins: propellers of cell function. *Trends Cell Biol* 10: 17-24.
 28. Kobayashi M, Yamamoto M (2005) Molecular mechanisms activating the Nrf2-Keap1 pathway of antioxidant gene regulation. *Antioxid Redox Signal* 7: 385-394.
 29. Prag S, Collett GDM, Adams JC (2004) Molecular analysis of muskelin identifies a conserved discoidin-like domain that contributes to protein self-association. *Biochem J* 381: 547-559.
 30. Kobayashi N, Yang J, Ueda A, Suzuki T, Tomaru K, et al. (2007) RanBPM, muskelin, p48EMLP, p44CTLH and ARMC8 and ARMC8 are components of the CTLH complex. *Gene* 396: 236-247.
 31. Zhou Z, Du X, Cai Z, Song X, Zhang H, et al. (2012) Structure of Sad1-UNC84 homology (SUN) domain defines features of molecular bridge in nuclear envelope. *J Biol Chem* 287: 5317-5326.