



Helicoverpa-inducible Thioredoxin h from *Cicer arietinum*: structural modeling and potential targets

Archana Singh^{a,1}, Chetna Tyagi^{b,1}, Onkar Nath^c, Indrakant K. Singh^{c,d,*}

^a Department of Botany, Hans Raj College, University of Delhi, Delhi, 110007, India

^b Department of Microbiology, University of Szeged, Közép fasor 52., 6726, Szeged, Hungary

^c Molecular Biology Research Laboratory, Deshbandhu College, University of Delhi, Kalkaji, New Delhi, 110019, India

^d Department of Entomology, University of Kentucky, S-225 AG. Science, North, Lexington, KY 40546-0001, United States

ARTICLE INFO

Article history:

Received 1 September 2017

Received in revised form 9 December 2017

Accepted 12 December 2017

Available online 17 December 2017

Keywords:

Herbivory

Chickpea

Thioredoxin h

Redox regulation

Comparative modeling

Docking

ABSTRACT

Thioredoxins are small and universal proteins, which are involved in the cell redox regulation. In plants, they participate in a broad range of biochemical processes like self-incompatibility, seed germination, pathogen & pest defense and oxidative stress tolerance. The h-type of thioredoxin (Trx-h) protein represents the largest Trx family. Herein, we characterized the *Helicoverpa* – inducible Trx h from an important legume, *Cicer arietinum*, CaHaTrx-h, 'CGFS' type Trxs, which encodes for a 113 amino acids long protein and possess characteristic motifs "FLKVDVDE" and "VVDFTASWCGPCRFIAPIL" and 73% sequence identity with AtTrx-h. Homology modeling and simulation of the target showed that the extended β -sheet regions remain stable during the simulation while the helical regions fluctuate between alpha and 3-₁₀ helical forms and highlights the flexibility of helix2-helix3 and terminal regions probably to accommodate an approaching protein target and facilitate their interaction. During the simulation, the structure exists in five energy minima clusters with biggest cluster size belonging to 20–25 ns time frames. PR-5 and Mannitol Dehydrogenase were nominated as potential targets and share close interaction with CaHaTrx-h via disulfide bond reduction. The study is an effort in the direction of understanding stress-related mechanisms in crop plants to overcome losses in agricultural yield.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Thioredoxins (Trxs) are a class of small disulfide reductases that regulate various redox-dependent processes [1–3]. Trxs were discovered in 1964 in an *E. coli* system as an electron donor of ribonucleotide reductase (an enzyme involved in DNA synthesis) that led to a great deal of interest in its multifarious roles [4]. They have multiple sub-cellular localizations such as cytosol, nucleus, mitochondria, and chloroplast or they may be released to the extracellular environment [5,6]. Out of six major categories of *Arabidopsis* Trx constituting h, f, m, o, x, and y based on different sub-cellular localization, Trx-h represents the largest structurally and biochemically complicated family [2,7]. Trxs play a significant role in the maintenance of cellular redox homeostasis as reductases [7–9], regulate programmed cell death via denitrosylation,

and also aid in defense responses [10–12]. The toxicity of reactive oxygen species (ROS) such as hydrogen peroxide is sequestered by Trx-h proteins through a thiol-disulfide exchange mechanism [13].

An active site dithiol, present in CXXC motif, partially exposed at the target protein surface is the key player in breaking of disulfide bonds in oxidized substrate proteins. These cysteine residues are oxidized and form a disulfide, which is converted back to the reduced form by NADPH-dependent thioredoxin reductase. There seems to be a minor but crucial structural difference between the reduced and oxidized forms of Trx and is mainly to do with the surface electronegativity of the dithiol based on its redox potential [14,15]. In addition, an example of their role in defense response has been evidenced in *Arabidopsis* and it has been shown that the expression of AtTrx-h3 and AtTrx-h5 is induced by certain pathogens and they contribute to systemic acquired resistance (SAR) [10,16]. We focused our investigation on Trx-h protein from *Cicer arietinum* to realize its function in plant defense.

Chickpea (*Cicer arietinum*) is an important food crop in the middle-eastern and southeastern parts of the world including India. *Helicoverpa armigera* infestation has been a major constraint for successful production of chickpea [17,18]. In our previous study,

* Corresponding author at: Molecular Biology Research Laboratory, Department of Zoology, Deshbandhu College, University of Delhi, Kalkaji, New Delhi, 110019, India.

E-mail address: iksingh@db.du.ac.in (I.K. Singh).

¹ These authors contributed equally.

we have shown that Trx homolog of chickpea namely, CaHaTrx-h changes its transcript level during herbivory and treatment with signaling compounds. Here, we report further characterization of CaHaTrx-h. The full-length sequence of CaHaTrx-h was retrieved and a computational study to elucidate its evolutionary relationships with other similar proteins was conducted. It is also crucial to understand its expression mechanism and to know the co-expressed genes and their regulation. A short molecular dynamics based simulation can also give major insight into dynamics of its structure, which is otherwise, well known for other model species like *Arabidopsis*. To confirm that it interact and regulates other proteins, two well-known targets of AtTrx-h were selected. These targets have been known to be implicated in plant defense responses and thus, can be correlated with our experimental results. This study throws light on the structural details, sub-cellular location, and interaction dynamics of CaHa-Trx-h. The use of time and cost-effective computational techniques is a growing necessity to overcome the vast search space required to understand molecular mechanisms. This study is a concerted effort to understand structure-function relationships and expression dynamics of CaHa-Trx-h protein. Characterizing important proteins implicated in defense responses of an agriculturally important plant species is the need of the hour.

2. Materials and methods

2.1. Experimental

2.1.1. Plant and insect growth conditions

Chickpea seeds (*C. arietinum* L.; Pusa-362) were procured from the Indian Agricultural Research Institute, New Delhi, India, were sown in pots containing autoclaved potting soil mixture (peat compost and vermiculite; 1:1 v/v). Plants were grown for 4 weeks in a greenhouse with 16/8 h light/dark cycle at 22–25 °C, 50–60% relative humidity (RH), and watered regularly during cultivation. Larvae of *H. armigera* were reared in the laboratory at 25 °C and 65–70% relative humidity (RH) on a 14/10 h light/dark cycle. The larvae were fed on an artificial diet as described by Armes et al. (1992) [19]. The freshly moulted fifth-instar larvae were starved over-night before releasing them on the plants.

2.1.2. Plant treatments

Insect infestation was achieved by the release of fifth-instar *H. armigera* larvae on 4-week-old chickpea plants (one larva per plant) and allowed to feed at 25 °C until 15–20% of the leaf area was consumed. Larvae were then removed, and the entire shoot was harvested and stored at –80 °C after quick freezing in liquid nitrogen. To mimic insect infestation, leaves were wounded with a punch machine (hole diameter 44.5 mm) (at 25 °C & 65–70% RH). Plants were subsequently harvested. For treatments involving exogenous signaling molecules, equal volumes of aqueous solutions of MeJA (100 μM; Aldrich, St Louis, MO, USA), SA (5 mM; Sigma, St Louis, MO, USA), and ethephon (50 μM, 2-chloroethanephosphonic acid, Sigma, St Louis, MO, USA) were sprayed onto chickpea plants according to published procedures [20]. The plants were then harvested and stored at –80 °C after quick freezing in liquid nitrogen.

2.1.3. Total RNA isolation and cDNA synthesis

Total RNA was isolated using RNeasy RNA isolation kit (Qiagen) following the manufacturer's protocol. The absorbance of RNA samples was determined using smart spectrophotometer (Biorad). RNA quality was measured on 1.2% formaldehyde agarose gel and by the optical density (OD) ratios at 260 nm/280 nm. In order to eliminate any genomic DNA contamination, total RNA of each samples were treated with DNase I (Qiagen). RNA was quantified using

spectrophotometer and equal quantity of each RNA sample was taken for cDNA synthesis. First strand cDNAs were synthesized using RevertAid cDNA synthesis kit (Thermo Scientific) following manufacturer's protocol.

2.1.4. Quantitative RT-PCR (qRT-PCR) and statistical analysis

For real time PCR, the cDNA was diluted 10 times and qRT-PCR was performed in triplicates in ABI StepOne Real-time (Applied biosystems) using SYBR Green Master Mix (Thermo Scientific) and gene-specific primers. Primers for qRT-PCR analysis were designed using the primer3 v.0.4.0 software [21,22] from unique regions. Amplicon lengths were kept 50–150 bases for optimal PCR efficiency. Each qRT-PCR reaction was performed in 10 μl reaction volume containing appropriately diluted cDNA as template, 225 nM of each forward and reverse primer, and 2 X Power SYBR Green PCR master mix. Thermal cycling conditions were 95 °C for 2 min followed by 40 cycles of 95 °C for 15 s, 60 °C for 1 min. The relative quantification method ($\Delta\Delta$ -CT) was used to evaluate quantitative variation between the replicates examined. The amplification of actin 2 was used as an endogenous control for normalization. Statistical significance of the differential expression patterns between treatments was determined using the Student's *t*-test. For considering a gene as *Helicoverpa*- or wounding-responsive or signaling compounds-responsive, the criterion of minimum 2-fold expression change (at least at one time point) with P-value, 0.05 was applied.

2.1.5. Expression patterns of caHa-Trx-h in different tissue types

For expression analysis of CaHaTrx-h in different tissues and organs during development, expression data available for the corresponding mRNA (TC04611) were retrieved from the Chickpea Transcriptome Database (CTDB) (<http://www.nipgr.res.in/ctdb.html>) and a graph corresponding to that data has been generated.

2.2. In silico analysis

2.2.1. Sequence alignments and phylogenetic analysis

The full protein sequence was obtained by translating the mRNA sequence with accession number XM_004500914.1 and is referred as CaHa-Trx-h (*Cicer arietinum* plant cDNA) throughout the manuscript. The analysis of CaHa-Trx-h protein was performed by ProtParam tool of EXPASY server [23]. The homology search was carried out using BLASTp program of NCBI [24] and sequence alignment was performed using MUSCLE software (v3.8.31) [25]. Web based tools such as TargetPv1.1 [26], MitoProt [27], and YLoc [28,29] were used to predict its putative cell localization. Prodist program of PHYLIP version 3.6 a3 [30] was used for studying phylogenetic relationships. The unrooted tree was made using neighbor-joining (NJ) method and plotted using DrawTree version 3.695 of PHYLIP.

2.2.2. Understanding functional interactive partners of thioredoxin h protein from *A. thaliana*

After ascertaining the similarity of CaHa-Trx-h protein with Thioredoxin h protein of *A. thaliana*, it is easier to assume that CaHa-Trx-h must show similarity in function and protein-protein interactions. As both seem to be part of the same machinery, they must elicit the same response to stress in terms of protein expression and co-regulation. To delve further we studied protein-protein association networks of *A. thaliana* Trx-h protein using the STRING database, version 10.5 [31]. Apart from already available experimental data, knowledge of known pathways and complexes from other databases, it includes data from co-expression analysis, literature search, gene orthology based knowledge transfer, and detection of shared selective signals across genomes.

Table 1
Sequences similar to CaHa-Trx-h along with their localization and active site.

Protein entry code	Organism name	Protein length	Putative localization	Active site sequence
CAA78462.1	<i>Arabidopsis thaliana</i>	114	Cytoplasm	CGFS
AAN76509.1	<i>Brassica rapa subsp. Oleifera</i>	133	Cytoplasm	CGFS
CAA55399.1	<i>Chlamydomonas reinhardtii</i>	113	Cytoplasm	CGFS
AAP33009.1	<i>Citrus x paradise</i>	123	Cytoplasm	CGFS
BAE16559.1	<i>Codonopsis lanceolata</i>	124	Chloroplast	CGFS
BAC21264.1	<i>Cucurbita maxima</i>	120	Cytoplasm	CGFS
ABB53600.1	<i>Eucalyptus grandis</i>	117	Chloroplast	CGFS
EKX51767.1	<i>Guillardia theta CCMP2712</i>	187	Extracellular	CGFS
AAU93947.1	<i>Helico sporidium sp. ex Simulium jonesi</i>	112	Cytoplasm	CGFS
AAD33596.1	<i>Hevea brasiliensis</i>	125	Chloroplast	CGFS
ACS96439.1	<i>Jatropha curcas</i>	118	Cytoplasm	CGFS
AAO16555.1	<i>Leymus chinensis</i>	131	Chloroplast	CGFS
AAZ32865.1	<i>Medicago sativa</i>	117	Cytoplasm	CGFS
AAZ42864.1	<i>Nicotiana glauca</i>	152	Cytoplasm	CGFS
AFP49340.1	<i>Olea europaea</i>	123	Cytoplasm	CGFS
AAB51522.1	<i>Oryza sativa Indica Group</i>	122	Cytoplasm	CGFS
EEC45257.1	<i>Phaeodactylum tricorutum CCAP 1055/1</i>	165	Extracellular	CGFS
CAC36986.1	<i>Pisum sativum</i>	120	Chloroplast	CGFS
AAL90749.1	<i>Populus tremula x Populus tremuloides</i>	143	Cytoplasm	CGFS
AAL26915.1	<i>Prunus persica</i>	136	Cytoplasm	CGFS
EMS21648.1	<i>Rhodospiridium toruloides NP11</i>	241	Cytoplasm	CGFS
AEM42995.1	<i>Siraitia grosvenorii</i>	121	Cytoplasm	CGFS
CAB96931.1	<i>Triticum aestivum</i>	125	Cytoplasm	CGFS
ABX79345.1	<i>Vitis vinifera</i>	126	Chloroplast	CGFS

2.2.3. Homology modeling and validation

The three-dimensional structure of CaHa-Trx-h was systematically elucidated by online homology-based structure prediction server ModWeb version r189 [32]. The structural data is then stored in ModBase [33]. It works by searching for the functionally and structurally closest protein template. A number of models are calculated if more than one template is available in the database. However, only one model was predicted in case of CaHa-Trx-h. The template search is carried out using PSI-BLAST or position-specific iterated BLAST search and is also used for fold assignment. The best structure is selected based on ModPipe Quality score (MPQS) and sequence identity which is discussed in the results section. The best scoring model was automatically selected. The presence of conserved structure was studied using NCBI Conserved Domain analysis [34] on modeled CaHa-Trx-h protein. An additional validation was done by the PSVS [35] server that includes a number of other algorithms namely, DSSP (secondary structure prediction) [36,37], MolProbity [38], PDB validation, Verify3D [39], and PROCHECK 3.5.4 [40]. An additional energy-based validation was also run on Prosa server [41].

2.2.4. A short molecular dynamics based simulation of modeled CaHa-Trx-h

The three-dimensional structure of CaHa-Trx-h was simulated for a short duration to delve into its dynamic properties. AmberTools16 [42] were used for carrying out this simulation using amberff14SBB force fields. Before the production run, the protein was energy minimized to clean up the structure and stabilize for MD simulations. The minimization was carried out for 1000 steps; 100 steps of steepest descent which later is switched to conjugate gradient method. The pairwise Generalized Born Solvent method was used. No cutoff was used to run a non-periodic simulation. A convergence criterion (of energy gradient) value of 0.01 was applied. The energy minimized coordinates were then put up for production run using Born implicit solvent model and no periodicity. The output was written to file every 500 steps. Langevin thermostat was used to regulate the temperature of the system at 300 K with initial and final temperatures to 300 K. The simulation ran for 25 ns with a 2 fs timestep. The SHAKE algorithm was used to constrain bonds involving hydrogen. This short simulation will help us to understand local motions in the overall topology and find out

highly fluctuating regions. The trajectory was visualized by VMD and its secondary structure calculated using its “Timeline” module. The Principal Component Analysis (PCA) was done using the *cpptraj* [43] module of AmberTools16. The porcupine plots were created using VMDs [44] Nmviz utility. The PCA plot was made using MDtraj [45] tools accessed through iJupyter notebook.

2.2.5. Mechanism of action through putative Arabidopsis thioredoxin H targets

To ascertain the functional aspects of the newly reported protein, its three-dimensional structure was docked with two known targets of Thioredoxin h in *Arabidopsis*. These two targets are mannitol dehydrogenase (ManD) and Pathogenesis-related 5 protein (thaumatin). During pathogen attack, plants exhibit excessive oxidative stress, which is dodged by some fungal pathogens due to the production of mannitol, a well-known ROS quencher. In turn, plants produce mannitol dehydrogenase to catabolize mannitol produced by fungal pathogens and thereby making it susceptible to damage due to reactive oxygen species (ROS) [46]. The latter protein, as the name suggests, is directly implicated in plant defense and belong to the thaumatin family. Both of them have been listed as putative targets in Marchand et al., 2004 [46] among many others but we chose these given their relevance in plant defense systems. Also, the other similar targets are not so well-known and hence an appropriate homologous template wasn't found for structure modeling.

Since their experimentally determined three-dimensional structures were not available (for *C. arietinum*), we resorted to homology-based modeling using Modbase, an online server for Modeller. With 71% and 50% identity shared with corresponding templates for ManD and thaumatin, respectively, the structures were obtained and analyzed. As a highly homologous template for Thaumatin wasn't found, we also modeled its structure using I-TASSER, an *ab initio* modeling server. The protein models thus obtained were validated using PSVS as described before. These proteins were docked with CaHa-Trx-h using Haddock server [47]. The active site residues were chosen to be cysteines from all three proteins. Thioredoxins have a conserved site Cys40-Cys43 that takes part in reducing target proteins. The average bond distance of a disulfide bond in a protein is 2.05 Å, just 0.5 Å more than a C–C bond. The S atoms of a disulfide bond in a target protein are reduced

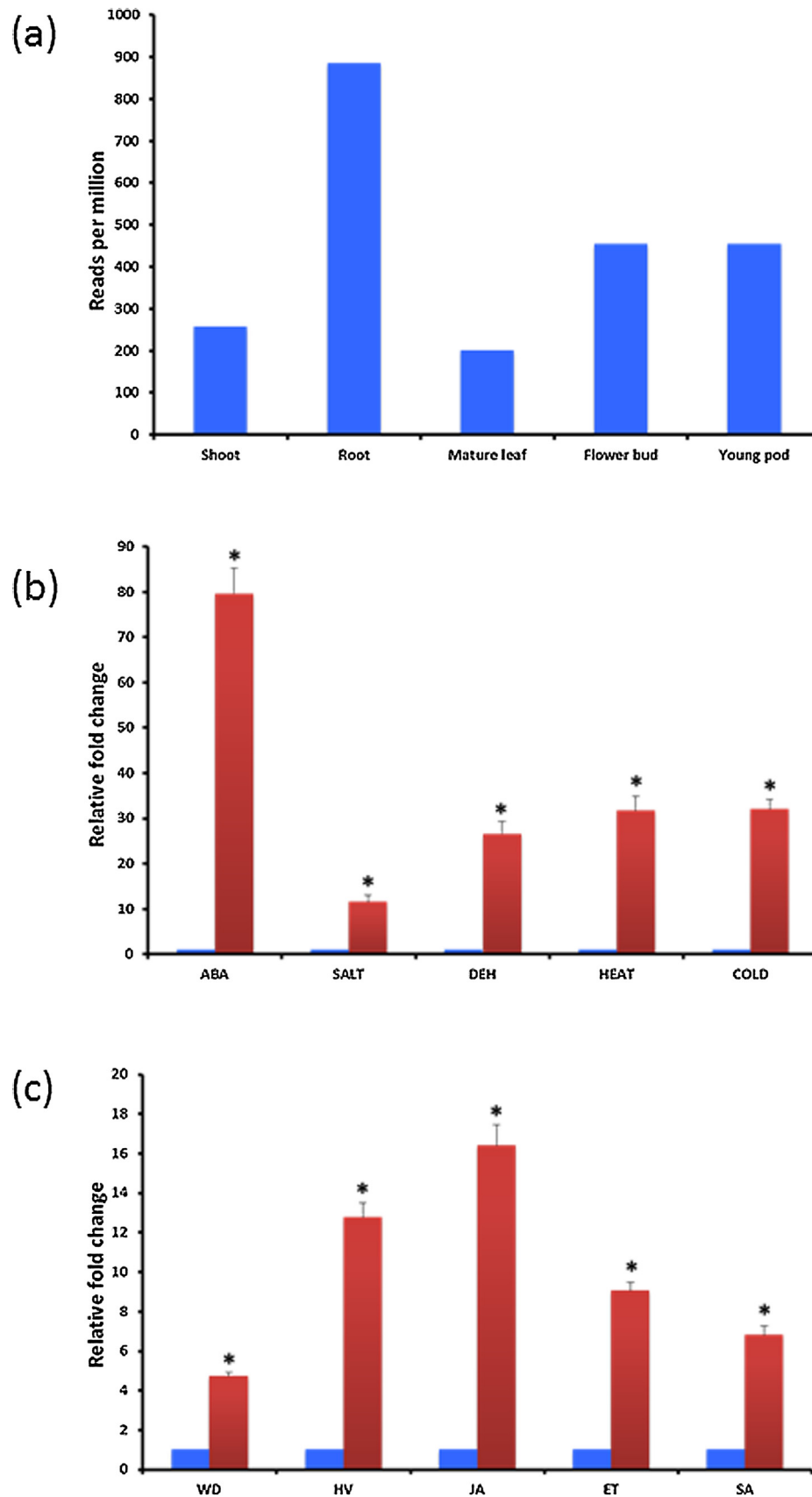


Fig. 1. (a) Expression Analysis of CaHaPR-4A in different tissue types. (Expression data has been retrieved from CTDB); (b) Expression analysis of Trx-h during different types of abiotic stresses and ABA, (c) Expression analysis of CaHaPR-4A during *Helicoverpa*-infestation (HV), Wounding (WD), Salicylic Acid (SA), Methyl Jasmonate (MeJa) and Ethephon (ET) treatments.

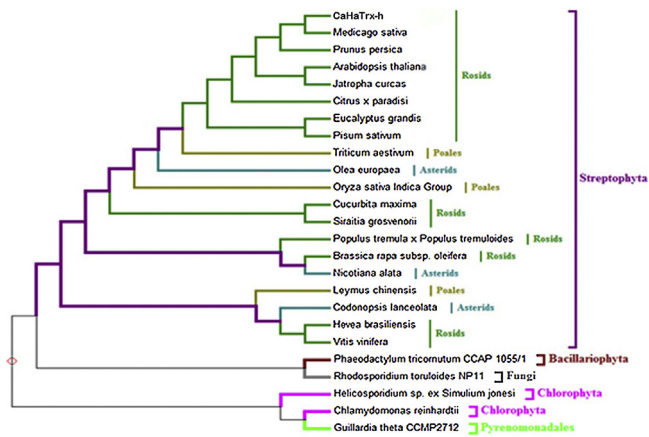


Fig. 2. Phylogenetic relationships of selected Thioredoxin sequences from different plant species.

and the bond is broken. Two cysteines forming a disulfide bond is collectively called a cystine residue.

3. Result and discussion

3.1. In silico analysis of caHa-Trx-h

The partial clone of CaHa-Trx-h was isolated from *Helicoverpa*-infested Chickpea Library. The full-length mRNA sequence and CDS was utilized for further analysis. This gene has an ORF of 342 bp and the deduced protein sequence of 113 amino acids. CaHa-Trx-h has the molecular weight of 12.39 kDa and theoretical pI of 5.67. The grand average hydropathicity (GRAVY) value is 0.033 and the instability index is 13.05, which signify the protein's stability.

CaHa-Trx-h showed highest homology with 1XFL chain A (Solution Structure Of Thioredoxin H1) of *Arabidopsis thaliana*, 1TI3 chain A (Solution Structure Of The Thioredoxin H1 From Poplar, A Cppc Active Site Variant) of *Populus tremula*, 1WMJ chain A (Solution Structure Of Thioredoxin Type H From) of *Oryza sativa*, 2VLT chain A (Crystal Structure Of Barley Thioredoxin H Isoform 2 In The Oxidized State) of *Hordeum vulgare*. The respective protein entry code, protein length and putative localization have been shown in Table 1. The CaHa-Trx-h was predicted to have a cytoplasmic localization. According to the classification of plant H-type subfamily 'FLKVDVDE' motif has been characterized. Another 'VVDFTASWCG-PCRFIAPIL' region was identified as the motif region in CaHa-Trx-h.

3.2. Expression analysis of caHa-Trx-h

Expression pattern of CaHa-Trx-h mRNA was analyzed using qPCR (Fig. 1). In the whole seedling, moderate levels of CaHa-Trx-h mRNA expression can be detected in normal growth conditions, indicating its requirement in the normal developmental process (Fig. 1a). The steady-state transcript of CaHa-Trx-h quickly reached to a higher level (twelve fold) within three hours of *Helicoverpa*-feeding. During mechanical wounding, CaHa-Trx-h transcript level increased (up to four fold) but not up to that extent as that of insect feeding. Signaling compounds (JA, SA and Et) play the crucial role in plant defense against pest and pathogen. Therefore, we studied the expression level of CaHa-Trx-h mRNA after treatment with these compounds. Accumulation of CaHa-Trx-h transcript was observed at greater level (sixteen fold) during JA-treatment and at a moderate level during Et-treatment (eight fold) and SA-treatment (six fold) (Fig. 1b). The critical role of thioredoxins during biotic stress has been stated [10] and it was shown that Trxs catalyzes the con-

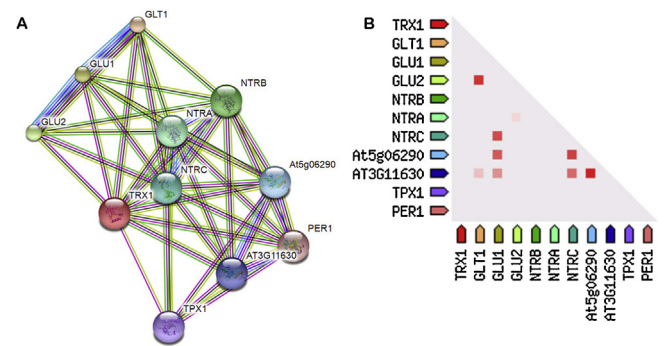


Fig. 3. (a) A network of co-expressed and regulated proteins in relation to *Arabidopsis thaliana* TRX-H protein, (b) A diagonal matrix describing the confidence score of co-expression of two proteins.

version of complex NPR1 to monomer [48,49], which is required for plant defense.

The involvement of thioredoxins is well documented during different types of abiotic stresses [50,51,52]. Considering that, we examined the change in its expression level during ABA-treatment, high salt, dehydration, heat-stress and cold-stress and observed that mRNA level increases thirty-fold during heat-stress and cold stress but to a lesser extent during dehydration (25 fold) and salt stress (10 fold). The CaHa-Trx-h transcript accumulation level was found to be at the maximum during ABA-treatment (Fig. 1c). These results clearly indicate the potential role of CaHa-Trx-h during abiotic stress. There are many reports suggesting the role of thioredoxins during abiotic stress. Over-expression of AtTrx-h3 confers heat resistance due to its chaperone function [50,51]. Similarly, OsTRXh1 was witnessed to be induced by salt stress and cold stress [52].

3.3. Phylogenetic analysis of caHa-Trx-h protein

To investigate phylogenetic relationships of CaHa-Trx-h with proteins of other plants, algae and other eukaryotes, an unrooted phylogenetic tree was constructed. The tree was sub-divided into five different clades: Streptophyta, Bacillariophyta, Fungi, Chlorophyta and pyrenomonadales (Fig. 2). The largest clade including CaHa-Trx-h of *Cicer arietinum*, Streptophyta comprises of Dicot and few Monocot plants like *Arabidopsis thaliana*, *Brassica rapa subsp. Oleifera*, *Citrus x paradise*, *Codonopsis lanceolata*, *Cucurbita maxima*, *Eucalyptus grandis*, *Hevea brasiliensis*, *Jatropha curcas*, *Leymus chinensis*, *Medicago sativa*, *Nicotiana glauca*, *Olea europaea*, *Oryza sativa Indica Group*, *Pisum sativum*, *Populus tremula x Populus tremuloides*, *Prunus persica*, *Siraiaia grosvenorii*, *Triticum aestivum*, *Vitis vinifera*. Sequences from *Rhodosporidium toruloides* NP11 and *Guillardia theta* CCMP2712 showed 42.94% and 59.50% divergence with the protein, respectively. Thus, both sequences were placed distantly from CaHa-Trx-h in the phylogenetic tree (Fig. 2). The close similarity of CaHa-Trx-H with *Medicago sativa* is not surprising (bootstrap value 97 out of 100) as both belong to the Fabaceae family while its similarity with peach (*Prunus persica*- bootstrap of only 28), a deciduous tree is quite opposite. The next clade contains *A. thaliana* and *J. curcas* (bootstrap 51), which are of high importance as model plants. This close similarity can be exploited to model thioredoxin systems in an *A. thaliana* system to be applied for crop plants like *C. arietinum*. *Citrus*, *Eucalyptus* and *Pisum sativum* are the rest three members of the closed clade.

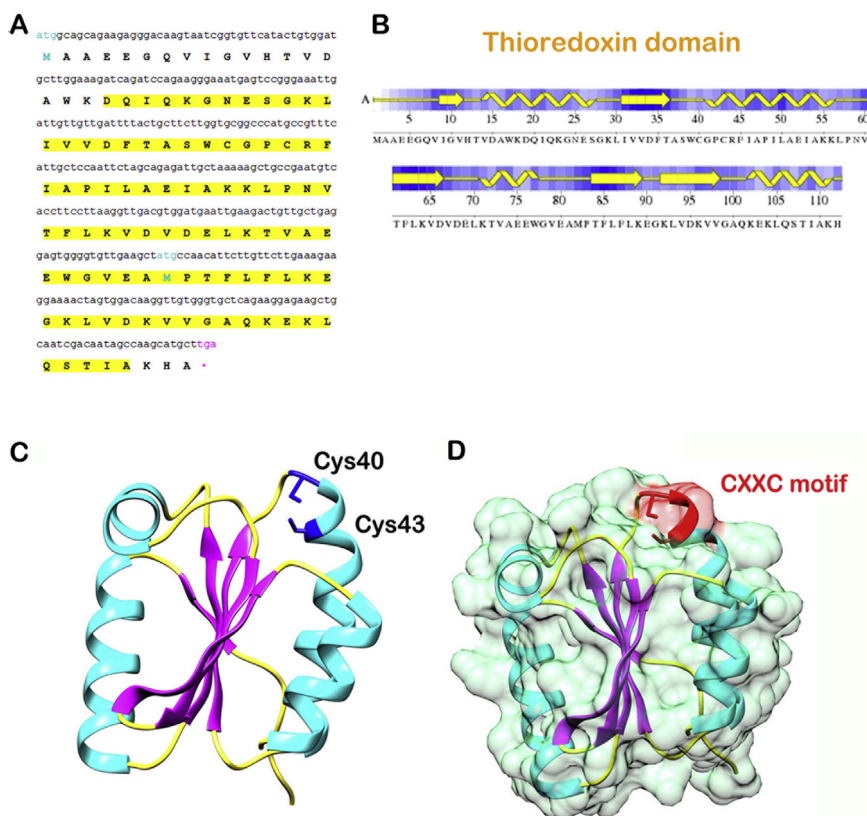


Fig. 4. (a) The translated sequence of predicted CaHa-Trx-h (primary structure), (b) secondary structure of the sequence, (c) The three-dimensional structure of CaHa-Trx-h protein with highlighted secondary structure elements, Cys40 and Cys43 are marked in blue, (d) the surface view of protein with active site CXXC motif present at the surface. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.4. Perspective from the closest relative: *A. thaliana* thioredoxin h [PDB: 1xf1]

To understand the protein–protein interaction network of CaHa-Trx-h, a great insight can be gained by observing *A. thaliana* Trx1 network. A preliminary search against the database shows its direct interaction with three proteins: NADPH-dependent thioredoxin reductases (NTRC, NTRA, and NTRB), glutamate synthases (GLT, GLU1, and GLU2), and periredoxins (periredoxin-2b TPX1, 1 Cys periredoxin PER1, and 2 Cys periredoxins AT3G11 and At5g06290). A graphical network is shown in Fig. 3a. Out of these, the structures of glutamate synthases are unknown as depicted by smaller spheres. Homologous relationships can be observed (as depicted by light blue connections) amongst each group of the network; all glutamate synthases, NTRA with NTRB and NTRC, and AT3G11 with At5g06290 (both are 2 Cys proteins). The black colored connections depict co-expression and all except TRX1 with TPX1 are shown to be co-expressed. This is a strong indication of coordinated stress-regulatory mechanisms. It was also suggested that TRX1 shares gene neighborhood with NTRs and periredoxins except TPX1 depicted by green connections. The other connections signify that the data was obtained from both experimental and curated databases. Gene co-expression in *A. thaliana* is described in detail in Fig. 3b. The intensity of red colored squares on the triangular matrix signifies the confidence that two proteins are functionally associated. TRX1 does not show co-expression with any other protein but other proteins can be seen showing strong co-expression. Glutamate synthases 1 and 2, 2 Cys periredoxins AT3G11 and At5g06290, and NADPH-dependent thioredoxin reductase with AT3G11 show a high confidence score. The STRING data is available in Fig. 1 and Fig. 2 of Supplementary Information 1.

3.5. Modeling and analysis of caHa-Trx-h protein structure

The protein-family search for CaHa-Trx-h showed the presence of thioredoxin domain (19–110 amino acids- Fig. 4a and b) with an E-value of $5.07e-36$. A total of 107 residues were found to have conserved folding among which almost all residues were conserved in a stretch, thus, proving that only the terminal regions are flexible. Only one CaHa-Trx-h model was predicted with *A. thaliana* Thioredoxin h1 as the template [PDB: 1xf1] [53] owing to 73% identity. The E value (expect value <0.0001) of zero indicates the significance of alignment and homologous relationship as reported by NCBI PSI-BLAST. The GA341 (Model score) indicates the reliability of the model based on derived statistical potentials. A score of 1.00 (higher than 0.7) is considered the reliable i.e. probability of correct fold assignment is more than 95%. The MPQS is a composite score of sequence identity, coverage, e-value, z-Dope (Discrete Optimized Protein Energy; threshold <0) and GA341 scores whose value of >1.1 is considered reliable. The score for CaHa-Trx-h is 1.99915 and thus can be considered a reliable model. The z-DOPE value is -1.86 indicating a reliable structure prediction. The overall α -RMSD (Root-Mean-Square Deviation) is calculated to be 1.215 Å and the predicted native overlap (3.5 Å) is 0.952 Å.

The multiple sequence alignment showed maximum conservation in the region of $\beta 1$ - $\alpha 1$ - $\beta 2$ - $\alpha 2$ - $\beta 3$ - $\alpha 3$ - $\beta 3$ - $\beta 4$ - $\alpha 3$, which is very similar to thioredoxin family. In CaHa-Trx-h, the thioredoxin motif was found in the $\beta 2$ - $\alpha 2$ region (Fig. 4). It started at the start point of the $\beta 2$ and continued till the mid-region of $\alpha 2$. Beta sheets form the core region of the protein. Alpha helices lie on the surface of the protein. First three beta strands are parallel ($\beta 1$, $\beta 2$ and $\beta 3$) and the other two are antiparallel ($\beta 4$ and $\beta 5$). The structure ends with alpha helix and contains other helices between the sheets (Fig. 4c and d).

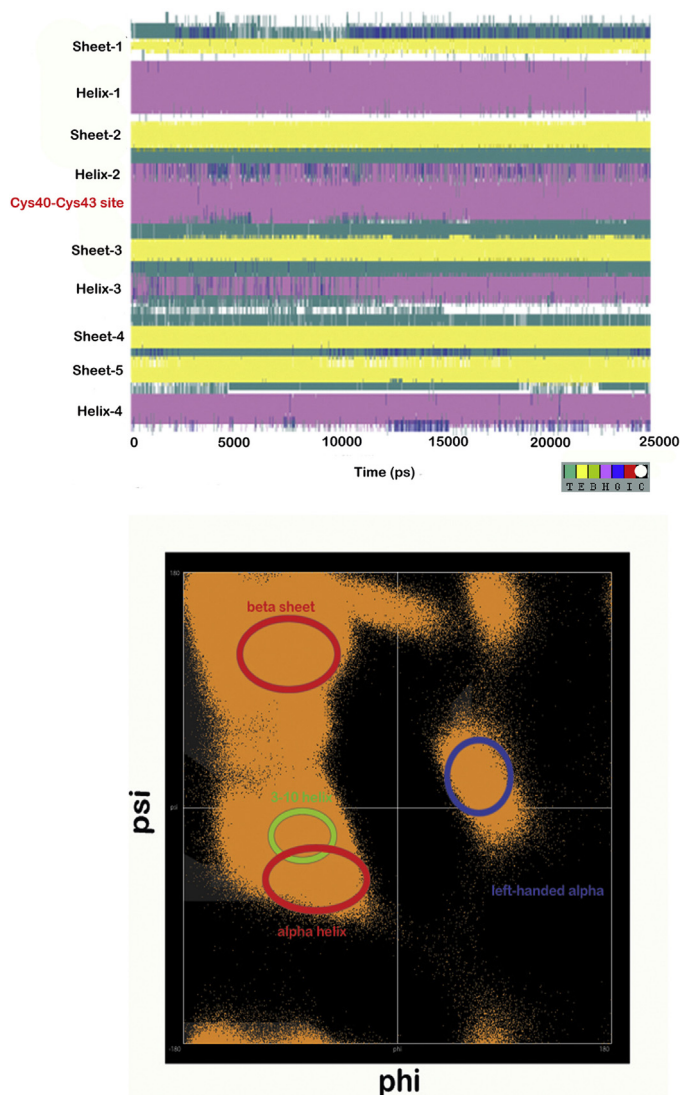


Fig. 5. (a) The evolution of secondary structure of CaHa-Trx-H with time during MD simulations. The active site lies in Helix 2, (b) Ramachandran plot of phi-psi values of every residue averaged for each time frame. The important secondary structural elements have been highlighted.

In one of the preliminary studies on structural and functional characters of Thioredoxins by Eklund et al. [54], Asp26, Ala29, Trp31, Cys32, Gly33, Pro34, Cys35, Asp61, Pro76, and Gly92 were listed to be conserved. The positions of these residues change slightly in CaHaTrx-h. The corresponding residues are Asp34, Ala37, Trp39, Cys40, Gly41, Pro42, Cys43, Asp68, Pro83, and Gly99. These residues form the surface around the active site and facilitate interaction. Other observed structurally important residue was Pro40 which corresponds to Pro48 which introduces a kink in $\alpha 2$ and positions the active site properly. Pro76 is considered important for maintaining the stability of the structure, which corresponds to Pro83 in CaHaTrx-h.

3.6. Validation of caHa-Trx-h model

The structure was validated using Ramachandra plot analysis (Fig. 3a-Supplementary file 1). The Procheck analysis showed 94.9% and 5.1% of residues in the favored and allowed regions respectively while no residue was detected as an outlier. Similar results were recorded for Richardson Lab's MolProbity with 99.1% and 0.1% of residues in the favored and allowed regions respectively. The

Global Quality scores from different modules were calculated as follows: Procheck G-factor (phi/psi) is 0.20, G factor (all dihedral angles) is -0.01 , and Verify3D score is 0.46. The G-factor indicates the goodness of the model. In other words, it's log-odds score based on the observed distribution of stereochemical parameters of proteins like the dihedral angles phi-psi and side-chain angles. These scores have been obtained by observing these parameters in 163 non-homologous structures obtained by X-ray crystallography. For a residue, a negative G-factor indicates a low-probability conformation and vice versa. Here, the mean G-score is good for phi-psi angles but not for side-chain angle parameters which is a limitation in case of computational prediction methods. The Verify3D score indicates the compatibility of a protein model (3D) with its amino acid sequence (1D) on the basis of the structural class assigned to each residue (alpha, beta, loop, polar, non-polar etc). A positive score of 0.46 (range $-1 \sim$ bad to $+1 \sim$ good) indicates an overall good 1D-3D compatibility. No close contacts were found within the range of 2.2 Å. The RMS deviation for bond angles and bond lengths was just 1.9° and 0.019 Å, respectively.

An additional run to validate the structure by Prosa yielded interesting results that provides an overall quality score and compares it with the native protein range. It takes only C-alpha atoms into account to make this comparison. The Z score of this calculation is -5.96 and lies within the range of experimentally determined structures (Fig. 3B-Supplementary file 1). The CaHa-Trx-h model does not deviate in terms of total energy of the structure. Another energy plot in Fig. 3B (Supplementary file 1) represents local model quality based on energies calculated for each residue. The positive values correspond to erroneous parts of the model. To avoid large fluctuations that may occur due to single residue plotting, a window of 10 and 40 residues is selected for accurate energy calculation. This explains a slight positive inclination of energy plot along a 10-residue window in Fig. 3B (Supplementary file 1) which is significantly reduced in the 40-residue window plot.

3.7. Analyzing the structure based on short molecular dynamics simulations

The CaHa-Trx-h predicted protein was simulated for a short time to study molecular fluctuation and dynamics taking place on a 25 ns timescale. After an initial energy minimization which stabilized the structure, actual production output was used to calculate the secondary structure evolution over time (Fig. 5a) and a Ramachandra plot was calculated based on average positions of dihedral angles phi and psi throughout the simulation (Fig. 5b). In Fig. 5a, the colors depict 6 different secondary structural elements like teal for turns (T), yellow for extended beta sheets (E), green for isolated bridge (B), magenta for alpha-helix (H), blue for 3–10 helix (G), red for pi-helix (I), and white for random coils (C). The secondary structure profile of CaHa-Trx-h shows a stable beta-sheet structure throughout the length of simulation indicating stability while the helix regions fluctuate to different conformations. A close overlap between alpha-helix and 3–10 helices can be seen for all helical regions. All helices are exposed to the surface and face stronger steric perturbation from the environment. These also form what is called as the interactive sites like the Cys40-Cys43 disulfide bond which is fairly exposed. The disulfide bond is surrounded by flexible loop regions which impart greater flexibility to the arrangement with a potential target. It also indicates towards a possible shift in structure due to varying redox states of CaHaTrx-h. In Fig. 5b, all important secondary structural elements are marked with circles. A left-handed helix formation also takes place along with few non-secondary structural regions populated by Proline and Glycine residues.

RMSF (root-mean-square- atomic fluctuation for each residue averaged over the trajectory) and RMSD graphs are shown in Fig. 6a

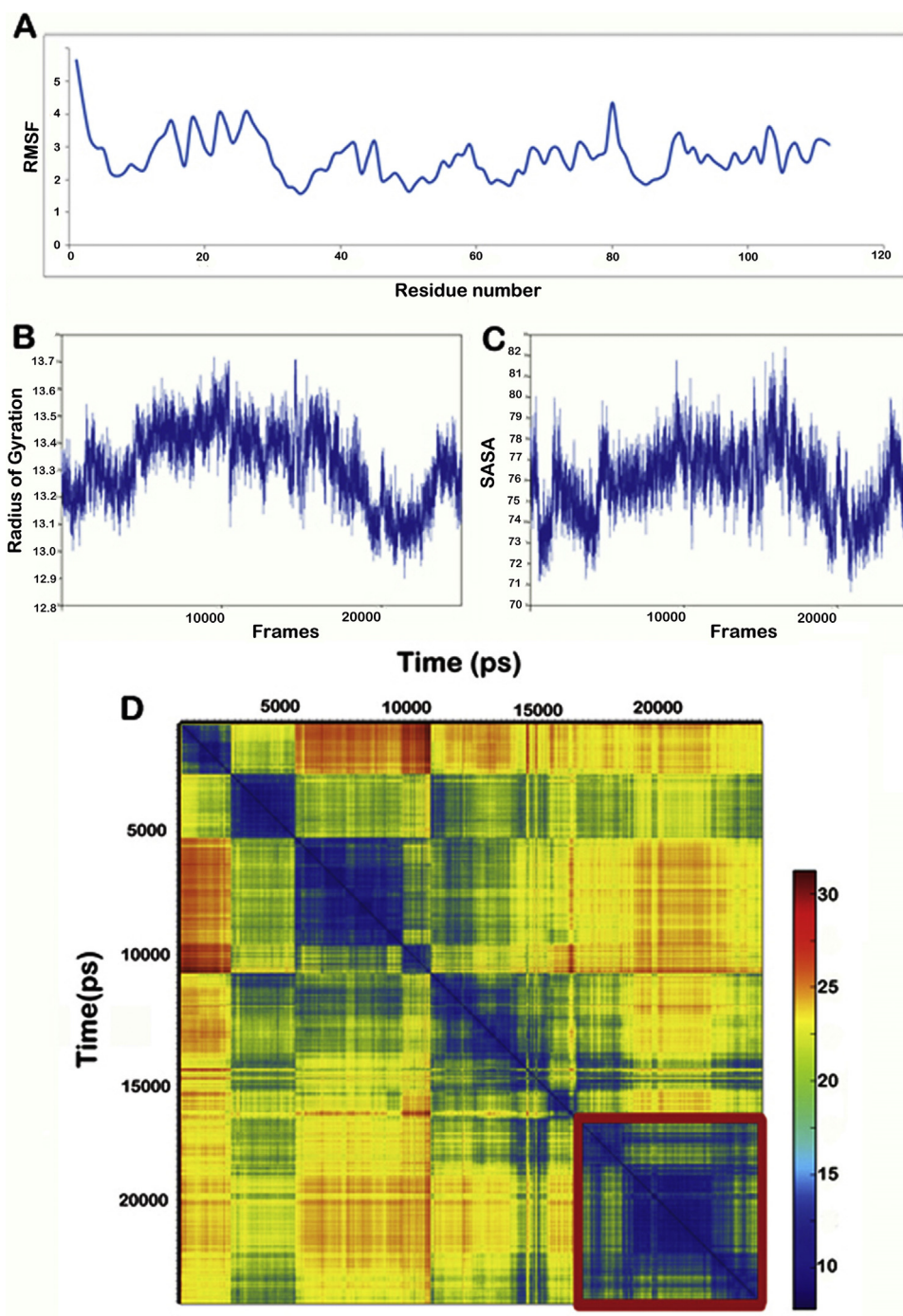


Fig. 6. (a) The root-mean-square-fluctuation of each residue along the protein chain, (b) radius of gyration along an imaginary axis depicting protein compactness, (c) the Solvent accessible surface area profile depicting the area in contact with the solvent during the simulation, and (d) 2-dimensional matrix of root-mean-square-deviation of backbone carbon atoms of all frames. The red square indicates the biggest cluster of low-RMSD structures formed towards end of simulation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and d, respectively. The RMSD graph is a 2-dimensional matrix of deviation calculated over time. The blue colors show low RMSD and thus, higher similarity and vice versa for regions depicted by yellow. The matrix shows that the structure exists in five states during the simulation. The largest cluster forms at the end of simulation indicating convergence.

The Radius of Gyration (RoG) and Solvent Accessible Surface Area (SASA) for the trajectory are shown in Fig. 6b and c, respectively. Both RoG and SASA plots signify that the protein structure was more compactly folded during the simulation, thereby,

decreasing the value of radius of molecule from an arbitrary axis that passes through the centre-of-mass of the structure. Decreasing SASA values also signify compact packing with less area exposed to solvent.

After an initial energy minimization which stabilized the structure, actual production output was studied using techniques like Normal Mode Analysis (NMA) [55] and Principal Component Analysis (PCA) [56].

The analyses show that the most fluctuating regions of the protein are the N- and C- terminals. The porcupine plots in Fig. 7a, and

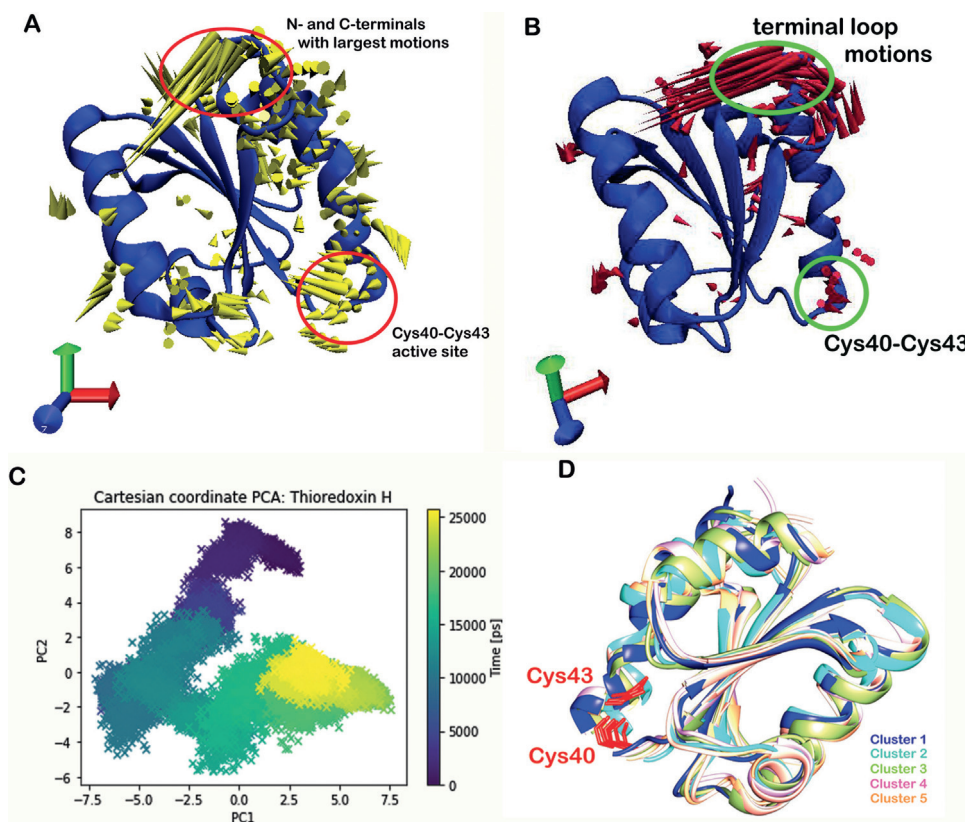


Fig. 7. (a and b) The two modes of motion depicted by colored porcupine needles. Most fluctuation can be seen in the surface elements and terminals. The active-site dithiols are also present at the surface and are highlighted with circles; (c) Principal component based clustering of various snapshots of the simulation trajectory based on respective Cartesian coordinates. The structure traverses a large search space during dynamic time evolution. The structure seems to exist in five minimum energy clusters, (d) A representative structure from each cluster is superimposed and the active site is marked.

b depict the two modes of motion. The direction and magnitude of motions is shown using varying sizes of the pointers. Large-sized pointers depict major fluctuations and vice versa. A cone extending from the C-alpha position of each residue shows the direction of the atom along the eigenvector. The terminals are almost never conserved and show high degrees of fluctuation (as depicted by the longer size of pointers) while the protein core is highly conserved and shows little or no fluctuation. The active site Cys40-Cys43 (Helix 2) was also observed with a large mode of motion. The direction of movement of helices 2 and 3 depict that these two regions move with respect to each other in an open-close manner to facilitate approaching target. The residues that form internal contact surface between secondary structural elements were observed by Eklund et al. [54], as Phe12, Val25, and Phe27 which corresponds to Val33 and Phe35 while the first Phenylalanine residue is not present.

A simple PCA scatter plot as shown in Fig. 7c depicts the spread of data (in this case the 3D cartesian coordinates of the system taken from all frames of the trajectory). To understand the changes undergone by the system, PCA is used to explain the variance in the coordinate space. The principal components (PCs as in the x- and y-axes) depict the modes of motion and the first PC represents the dominant motion and so on. A plot of PC1 versus PC2 shows trajectory frames as colored data-points (after dimensionality reduction). The color-bar depicts the time in ps. Note how the points graduate towards a completely different conformational subspace as a function of time. The most dominant motion takes place in the 10 ns time step while the second most dominant motion takes over after at about 20 ns. It also depicts that the structure exists in about 5 energy minima clusters and different representative structures can be attributed to each stage (Fig. 7d). It is understood that the actual

protein motion is depicted as the sum of all individual modes of motion. The significance lies in the fact that even on a nanosecond time scale; huge molecular fluctuations can occur and give rise to the plethora of protein-protein interactions and signaling. This shows that proteins are not rigid structures and apart from a stable core region, the solvent-exposed regions are constantly changing to interact and bind with targets or membranes.

A validation run of the average structure resulted in a lower Z-score of -5.3 from Prosa server, which signifies a structure of overall quality when compared with experimentally determined structures.

3.8. Modeling and validation of three-dimensional structures of ManD and thaumatin

As described earlier, due to the absence of experimentally determined structures of these target proteins for *C. arietinum*, they were modeled and validated using the same tools as for CaHa-Trx-h. The template chosen for ManD showed 71% identity and is an alcohol dehydrogenase from *Populus tremuloides*. The various scores for this model are as follows: E-value = 0; GA341 = 1.00; zDOPE = -1.13 ; RMSD = 2.077; Native contact overlap = 0.94. All the calculations indicate towards a reliable model and fold prediction. The other target Thaumatin was modeled using I-TASSER server because of low-identity template search. It works by taking into account multiple protein templates that share highly similar domains or motifs with the query. Many templates were used for this prediction: Thaumatin from *Thaumatococcus danielli* [2pe7, 2vu7, 3e3s, 2d8o], pathogenesis-related protein from *Nicotiana tabacum* [1aun], anti-fungal Zeatin from *Zea Mays* [1du5], *Prunus avium* allergen protein [2ahn- also selected as the Modweb tem-

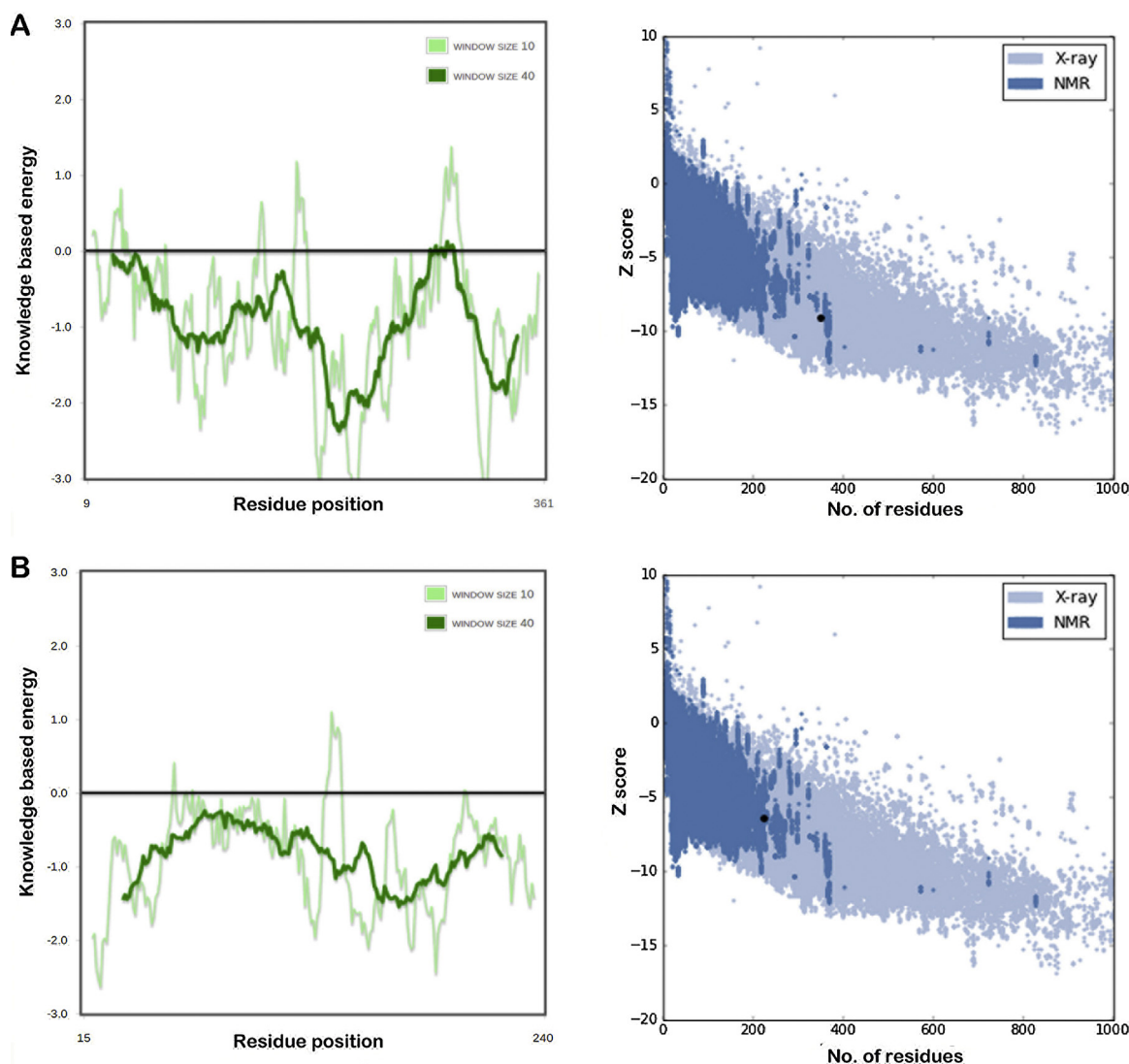


Fig. 8. Energy plot of the modeled protein in which positive values correspond to erroneous parts. The two lines depict two separate window length. The Z-score plot compares the Z score with those of experimentally known proteins, (a) Mannitol dehydrogenase, (b) Thaumatin.

plate]. The continuous fragments obtained by threading alignment are extracted and reassembled using replica-exchanged Monte Carlo simulations. SPICKER was used to cluster the generated decoy structures with low temperatures and top five cluster centroids are selected for generating full atomic models. Out of five highly probable structures resulting from this calculation, one with the highest C score and TM score is selected. The C-score is a measure of confidence in a predicted structure based on the template alignments. It ranges from -5 to 2 with a higher score depicting higher confidence in the model. The predicted thaumatin model has a C-score value of -1.38 , which is not very high but also not very low. Another such value is the TM score, which imparts a better understanding of the distance of the predicted structure from the native. Previous methods like root-mean-square-prediction (RMSD) can be ambiguous due to large impact of local errors even if the global topology is correct. A TM-score >0.5 indicates a model of correct topology and a TM-score <0.17 means a random similarity. The TM-score value of the modeled structure is 0.54 ± 0.15 . The thaumatin model was generated from a cluster with 3109 decoys and a cluster density of 0.0564 , which is the higher than all other clusters and signifies ample occurrence of selected structure over the course of simulation trajectory.

For validating these structures, the PSVS results are mentioned here. For ManD protein model, the MolProbity Ramachandran analysis predicted 96.9% of its residues falling in the favored regions and only 0.6% in the disallowed regions. The Global Quality scores from different modules were calculated as follows: Procheck G-factor (ϕ/ψ) is 0.00; G factor (all dihedral angles) is 0.03, and Verify3D score is 0.50, all of which indicate towards a reliable model. No close contacts were found within the range of 2.2 \AA . The RMS deviation for bond angles and bond lengths was just 2.1° and 0.018 \AA , respectively. For thaumatin model, MolProbity Ramachandran analysis predicted 96.0% of its residues falling in the favored regions and only 0.9% in the disallowed regions. The Global Quality scores from different modules were calculated as follows: Procheck G-factor (ϕ/ψ) is -0.36 , G factor (all dihedral angles) is -0.27 and the Verify3D score is 0.46, all of which indicate towards a reliable model. No close contacts were found within the range of 2.2 \AA . The RMS deviation for bond angles and bond lengths was just 2.3° and 0.020 \AA , respectively. Out of all models, predicted thaumatin structure has the most unreliable fold prediction. The N- and C-terminals show the highest fluctuation and less accuracy due to the absence of a highly homologous protein of known experimental structure. The two structures were also validated using the PROSA server

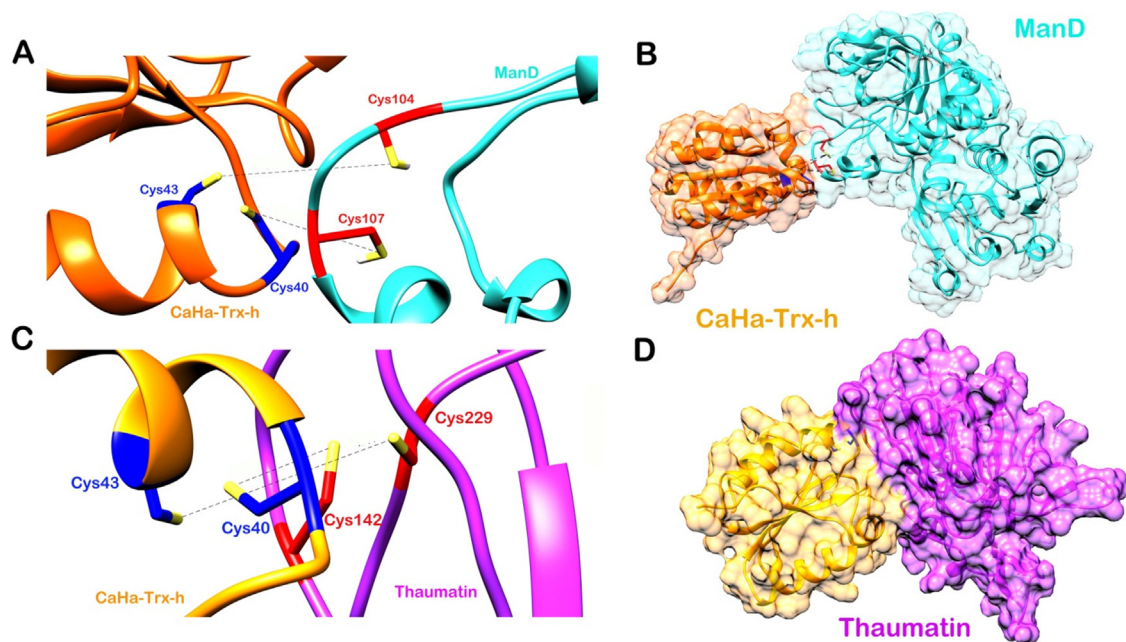


Fig. 9. (a) Cysteine residue interaction between CaHa-Trx-h and ManD target protein, (b) surface view of CaHa-Trx-h-ManD complex, (c) residue interaction between CaHa-Trx-h and thaumatin target protein, (d) surface view of CaHa-Trx-h-thaumatin complex.

and resulted in energetically stable conformations comparable to experimentally known structures (Fig. 8).

3.9. Identification of potential targets of caHaTrx-h and understanding its mechanism of function

To elucidate the mechanism of function of the modeled protein CaHa-Trx-h, its interaction pattern with two putative protein targets of AtTrx was studied. The putative targets described by Marchand et al., [48] implicated in defense mechanisms are (a) myrosinase, which is a glycoprotein and the glucosinolate-myrosinase system is a characteristic of family Brassicaceae, (b) apospory associated protein C. Due to the unavailability of further structural knowledge for *Arabidopsis* myrosinase and protein C, we decided not to select them for further studies. The selected protein targets are mannitol dehydrogenase (ManD) and thaumatin as identified for an *A. thaliana* system. The reduction mechanism of Trxs is a disulfide reducing system through a well-conserved disulfide active site (WCGPC). To validate that CaHa-Trx-h is indeed a Thioredoxin h from *C. arietinum*, we assume that it should interact with the selected protein targets and reduce them through its disulfide (C–C) bridge. Cys40–Cys43 disulfide bridge of Thioredoxin h has been selected as the active site based on previous studies. The disulfide bridge formed at the target protein surface is targeted to monitor this interaction. The HADDOCK score for CaHa-Trx-h–ManD interaction was -74.4 ± 4.0 . About 182 complex structures were predicted and grouped into 11 clusters. The top-ranked cluster has 41 complex structures with a 13.6 ± 0.2 RMSD from the overall lowest-energy structure. The energies were also calculated as follows, Van der Waals energy: -28.1 ± 2.9 , electrostatic energy: -162.3 ± 22.2 , desolvation energy: -14.1 ± 3.1 , restraints violation energy: 2.4 ± 0.54 . The buried surface area of the complex is 890.5 ± 113.5 . Lastly, the Z score value of -1.3 indicates how many standard deviations this cluster is located from the average in terms of the score; the more negative the better. The top-ranked structure from this cluster was selected for analysis. The interaction is depicted in Fig. 9a and b.

The HADDOCK score for CaHa-Trx-h–thaumatin interaction was -103.5 ± 7.4 . About 175 complex structures were predicted and

grouped into 10 clusters. The top-ranked cluster has 29 complex structures with a 0.8 ± 0.5 RMSD from the overall lowest-energy structure. The energies were also calculated as follows, Van der Waals energy: -49.2 ± 3.8 , electrostatic energy: -211.3 ± 47.0 , desolvation energy: -14.6 ± 3.6 , restraints violation energy: 26.3 ± 17.76 . The buried surface area of the complex is 1390.9 ± 30.4 . Lastly, the Z score value of this complex structure is -1.8 . The top-ranked structure from this cluster was selected for analysis. The interaction constitutes a direct linking of Cys40–Cys43 (CaHa-Trx-h) with Cys142–Cys229 (thaumatin) disulfide bridge (Fig. 9c and d).

Comparing the two clusters, CaHa-Trx-h–thaumatin complex seems more stable than the former with a better HADDOCK score and Z score. The cluster RMSD from the average is also very less in comparison to the former. All energy values also indicate a more stable CaHa-Trx-h–thaumatin complex. However, the restraint violation energy is higher than in CaHa-Trx-h–ManD complex, which indicates that not all restraints were fulfilled. This may also indicate a larger structural shift during interaction and disulfide bond breakage leading to bigger violations. Owing to a correlation observed between binding energy and buried surface area by Chen et al., 2013 [57], it is no surprise that CaHa-Trx-h–thaumatin complex is more stable than the former. Moreover, transcript of PR-5 shows induction during *Helicoverpa* infestation [24], which further strengthens our prediction that CaHa-Trx-h interacts with PR-5.

4. Conclusion

The necessity of present-day world to find sustainable and rational methods to fight against various plant diseases has led many of us to delve deeper into plant defense mechanisms. Thioredoxins are an important class of proteins that are ubiquitously found in the living world. Amongst partaking in other crucial biological processes, their most important task is to quench the ROS stress. Various proteins can exist either in oxidized or reduced form defining different states of activity. The oxidized proteins are reduced by Thioredoxins. Trx-h, from *Cicer arietinum*, became the subject of this study when it was found to be expressed during *Helicoverpa* infestation. In this study, CaHaTrx-h was studied for its expres-

sion levels during various abiotic stresses and under application of different plant hormones. Its three-dimensional structure was modeled based on homology and dynamically simulated to elucidate the fluctuating regions. The active site di-cysteine, present at the surface, was seen to be driving the interaction with PR-5 (thaumatin) and mannitol dehydrogenase. The target proteins also contain an active site di-cysteine in the oxidized form, which in turn, is reduced by the approaching Trx-h. An advanced knowledge of protein systems and their function is crucial to device new techniques for agricultural management. This study is a step for the use of experimental and computational techniques for fast and efficient proteomics research.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

The authors AS and IKS are thankful to University Grant Commission, MHRD, Govt. of India for post-doctoral fellowship and DST-SERB, Govt. of India for research grant. CT is thankful to Department of Microbiology, TTIK, University of Szeged for providing computational resources.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.ijbiomac.2017.12.079>.

References

- [1] A. Holmgren, Thioredoxin, *Ann. Rev. Biochem.* 54 (1) (1985) 237–271.
- [2] E. Gelhaye, N. Rouhier, N. Navrot, J. Jacquot, The plant thioredoxin system, *Cell. Mol. Life Sci. CMLS* 62 (1) (2005) 24–35.
- [3] Y. Meyer, J.P. Reichheld, F. Vignols, Thioredoxins in Arabidopsis and other plants, *Photosynth. Res.* 86 (3) (2005) 419–433.
- [4] E.C. Laurent, Enzymatic synthesis of deoxyribonucleotides IV. Isolation and characterization of thioredoxin, the hydrogen donor from *Escherichia coli* B, *J. Biol. Chem.* 239 (10) (1964) 3436–3444.
- [5] C. Laloi, N. Rayapuram, Y. Chartier, J.-M. Grienerberger, G. Bonnard, Y. Meyer, Identification and characterization of a mitochondrial thioredoxin system in plants, *Proc. Natl. Acad. Sci.* 98 (24) (2001) 14144–14149.
- [6] A.J. Serrato, J.L. Crespo, F.J. Florencio, F.J. Cejudo, Characterization of two thioredoxins h with predominant localization in the nucleus of aleurone and scutellum cells of germinating wheat seeds, *Plant Mol. Biol.* 46 (3) (2001) 361–371.
- [7] E. Gelhaye, N. Rouhier, J.-P. Jacquot, The thioredoxin h system of higher plants, *Plant Physiol. Biochem.* 42 (4) (2004) 265–271.
- [8] H.J. Atkinson, P.C. Babbitt, An atlas of the thioredoxin fold class reveals the complexity of function-enabling adaptations, *PLoS Comput. Biol.* 5 (10) (2009) e1000541.
- [9] A.J. Serrato, F.J. Cejudo, Type-h thioredoxins accumulate in the nucleus of developing wheat seed tissues suffering oxidative stress, *Planta* 217 (3) (2003) 392–399.
- [10] Y. Tada, S.H. Spoel, K. Pajerowska-Mukhtar, Z. Mou, J. Song, C. Wang, J. Zuo, X. Dong, Plant immunity requires conformational charges of NPR1 via S-nitrosylation and thioredoxins, *Science* 321 (5891) (2008) 952–956.
- [11] L. Sun, H. Ren, R. Liu, B. Li, T. Wu, F. Sun, H. Liu, X. Wang, H. Dong, An h-type thioredoxin functions in tobacco defense responses to two species of viruses and an abiotic oxidative stress, *Mol. Plant Microbe Interact.* 23 (11) (2010) 1470–1485.
- [12] S. Kneeshaw, S. Gelineau, Y. Tada, G.J. Loake, S.H. Spoel, Selective protein denitrosylation activity of thioredoxin-h5 modulates plant immunity, *Mol. Cell* 56 (1) (2014) 153–162.
- [13] M.-Y. Lee, K.-H. Shin, Y.-K. Kim, J.-Y. Suh, Y.-Y. Gu, M.-R. Kim, Y.-S. Hur, O. Son, J.-S. Kim, E. Song, Induction of thioredoxin is required for nodule development to reduce reactive oxygen species levels in soybean roots, *Plant Physiol.* 139 (4) (2005) 1881–1889.
- [14] A. Holmgren, Thioredoxin structure and mechanism: conformational changes on oxidation of the active-site sulfhydryls to a disulfide, *Structure* 3 (3) (1995) 239–243.
- [15] A. Holmgren, M. Bjornstedt, Thioredoxin and thioredoxin reductase, *Methods Enzymol.* 252 (1995) 199–208.
- [16] C. Laloi, D. Mestres-Ortega, Y. Marco, Y. Meyer, J.-P. Reichheld, The Arabidopsis cytosolic thioredoxin h5 gene induction by oxidative stress and its W-box-mediated response to pathogen elicitor, *Plant Physiol.* 134 (3) (2004) 1006–1016.
- [17] K. Singh, Chickpea (*Cicer arietinum* L.), *Field Crops Res.* 53 (1–3) (1997) 161–170.
- [18] T. Shanower, J. Romeis, E. Minja, Insect pests of pigeonpea and their management, *Annu. Rev. Entomol.* 44 (1) (1999) 77–96.
- [19] N.J. Armes, G. Bond, R. Cooter, The Laboratory Culture and Development of *Helicoverpa armigera*, *Nat. Resources Inst.*, 1992.
- [20] A. Singh, I.K. Singh, P.K. Verma, 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 (9) (2008) 2379–2392.
- [21] A. Untergasser, I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, S.G. Rozen, Primer3—new capabilities and interfaces, *Nucleic Acids Res.* 40 (15) (2012), e115–e115.
- [22] T. Koressaar, M. Remm, Enhancements and modifications of primer design program Primer3, *Bioinformatics* 23 (10) (2007) 1289–1291.
- [23] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel, A. Bairoch, ExPASy: the proteomics server for in-depth protein knowledge and analysis, *Nucleic Acids Res.* 31 (13) (2003) 3784–3788.
- [24] M. Johnson, I. Zaretskaya, Y. Raytselis, Y. Merezuk, S. McGinnis, T.L. Madden, NCBI BLAST: a better web interface, *Nucleic Acids Res.* 36 (suppl.2) (2008) W5–W9.
- [25] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.* 32 (5) (2004) 1792–1797.
- [26] O. Emanuelsson, H. Nielsen, S. Brunak, G. Von Heijne, Predicting subcellular localization of proteins based on their N-terminal amino acid sequence, *J. Mol. Biol.* 300 (4) (2000) 1005–1016.
- [27] M.G. Claros, P. Vincens, Computational method to predict mitochondrially imported proteins and their targeting sequences, *The FEBS J.* 241 (3) (1996) 779–786.
- [28] S. Briesemeister, J. Rahnenführer, O. Kohlbacher, YLoc—an interpretable web server for predicting subcellular localization, *Nucleic Acids Res.* 38 (suppl.2) (2010) W497–W502.
- [29] S. Briesemeister, J. Rahnenführer, O. Kohlbacher, Going from where to why—interpretable prediction of protein subcellular localization, *Bioinformatics* 26 (9) (2010) 1232–1238.
- [30] J. Feltsenstain, {PHYLLIP} (Phylogeny Inference Package) version 3.6 a3 (2002).
- [31] D. Szklarczyk, J.H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N.T. Doncheva, A. Roth, P. Bork, The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible, *Nucleic Acids Res.* 45 (D1) (2017) D362–D368.
- [32] A. Fiser, A. Sali, Modeller: generation and refinement of homology-based protein structure models, *Methods Enzymol.* 374 (2003) 461–491.
- [33] U. Pieper, B.M. Webb, D.T. Barkan, D. Schneidman-Duhovny, A. Schlessinger, H. Braberg, Z. Yang, E.C. Meng, E.F. Pettersen, C.C. Huang, ModBase, a database of annotated comparative protein structure models, and associated resources, *Nucleic Acids Res.* 39 (suppl.1) (2010) D465–D474.
- [34] A. Marchler-Bauer, Y. Bo, L. Han, J. He, C.J. Lanczycki, S. Lu, F. Chitsaz, M.K. Derbyshire, R.C. Geer, N.R. Gonzales, CDD/SPARCLE: functional classification of proteins via subfamily domain architectures, *Nucleic Acids Res.* 45 (D1) (2016) D200–D203.
- [35] A. Bhattacharya, R. Tejero, G.T. Montelione, Evaluating protein structures determined by structural genomics consortia Proteins, *Struct. Funct. Bioinf.* 66 (4) (2007) 778–795.
- [36] R.P. Joosten, T.A. Te Beek, E. Krieger, M.L. Hekkelman, R.W. Hooft, R. Schneider, C. Sander, G. Vriend, A series of PDB related databases for everyday needs, *Nucleic Acids Res.* 39 (suppl.1) (2010) D411–D419.
- [37] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers* 22 (12) (1983) 2577–2637.
- [38] V.B. Chen, W.B. Arendall, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr. Sect D Biol. Crystallogr.* 66 (1) (2010) 12–21.
- [39] R. Luthy, J.U. Bowie, D. Eisenberg, Assessment of protein models with three-dimensional profiles, *Nature* 356 (6364) (1992) 83.
- [40] Y.J. Huang, R. Powers, G.T. Montelione, Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics, *J. Am. Chem. Soc.* 127 (6) (2005) 1665–1674.
- [41] M. Wiederstein, M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, *Nucleic Acids Res.* 35 (suppl.2) (2007) W407–W410.
- [42] D.A. Case, R.M. Betz, D.S. Cerutti, T.E. Cheatham, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, J.R. Lee, S. LeGrand, B. Li, C. Lin, T. Luchko, R. Luo, M. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J.M. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao, P.A. Kollman, AMBER (2016) (2016).
- [43] D.R. Roe, T.E. Cheatham III, PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data, *J. Chem. Theory Comput.* 9 (7) (2013) 3084–3095.

- [44] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, *J. Mol. Graph.* 14 (1) (1996) 33–38.
- [45] R.T. McGibbon, K.A. Beauchamp, M.P. Harrigan, C. Klein, J.M. Swails, C.X. Hernández, C.R. Schwantes, L.-P. Wang, T.J. Lane, V.S. Pande, MDTraj A modern open library for the analysis of molecular dynamics trajectories, *Biophys. J.* 109 (8) (2015) 1528–1532.
- [46] C. Marchand, P. Le Maréchal, Y. Meyer, M. Miginiac-Maslow, E. Ilssakidis-Bourguet, P. Decottignies, New targets of Arabidopsis thioredoxins revealed by proteomic analysis, *Proteomics* 4 (9) (2004) 2696–2706.
- [47] G. Van Zundert, J. Rodrigues, M. Trellet, C. Schmitz, P. Kastiris, E. Karaca, A. Melquiond, M. van Dijk, S. De Vries, A. Bonvin, The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes, *J. Mol. Biol.* 428 (4) (2016) 720–725.
- [48] Y. Balmer, A. Koller, G. del Val, W. Manieri, P. Schürmann, B.B. Buchanan, Proteomics gives insight into the regulatory function of chloroplast thioredoxins, *Proc. Natl. Acad. Sci.* 100 (1) (2003) 370–375.
- [49] D. Yamazaki, K. Motohashi, T. Kasama, Y. Hara, T. Hisabori, Target proteins of the cytosolic thioredoxins in Arabidopsis thaliana, *Plant Cell Physiol.* 45 (1) (2004) 18–27.
- [50] J.R. Lee, S.S. Lee, H.H. Jang, Y.M. Lee, J.H. Park, S.-C. Park, J.C. Moon, S.K. Park, S.Y. Kim, S.Y. Lee, Heat-shock dependent oligomeric status alters the function of a plant-specific thioredoxin-like protein, AtTDX, *Proc. Natl. Acad. Sci.* 106 (14) (2009) 5978–5983.
- [51] S.K. Park, Y.J. Jung, J.R. Lee, Y.M. Lee, H.H. Jang, S.S. Lee, J.H. Park, S.Y. Kim, J.C. Moon, S.Y. Lee, Heat-shock and redox-dependent functional switching of an h-type Arabidopsis thioredoxin from a disulfide reductase to a molecular chaperone, *Plant Physiol.* 150 (2) (2009) 552–561.
- [52] G. Xie, H. Kato, K. Sasaki, R. Imai, A cold-induced thioredoxin h of rice, OsTrx23, negatively regulates kinase activities of OsMPK3 and OsMPK6 in vitro, *FEBS Lett.* 583 (17) (2009) 2734–2738.
- [53] F.C. Peterson, B.L. Lytle, S. Sampath, D. Vinarov, E. Tyler, M. Shahan, J.L. Markley, B.F. Volkman, Solution structure of thioredoxin h1 from Arabidopsis thaliana, *Protein Sci.* 14 (8) (2005) 2195–2200.
- [54] H. Eklund, F.K. Gleason, A. Holmgren, Structural and functional relations among thioredoxins of different species, *Proteins* 11 (1) (1991) 13–28.
- [55] D.A. Case, Normal mode analysis of protein dynamics, *Curr. Opin. Struct. Biol.* 4 (2) (1994) 285–290.
- [56] G.G. Maisuradze, A. Liwo, H.A. Scheraga, Principal component analysis for protein folding dynamics, *J. Mol. Biol.* 385 (1) (2009) 312–329.
- [57] J. Chen, N. Sawyer, L. Regan, Protein-protein interactions: general trends in the relationship between binding affinity and interfacial buried surface area, *Protein Sci.* 22 (4) (2013) 510–515.