A Conserved NAG motif is critical to the catalytic activity of galactinol synthase, a key regulatory enzyme of RFO biosynthesis

Prafull Salvi†, Bhavnesh Kumar, Nitin Uttam Kamble, Abhijit Hazra, and Manoj Majee*

National Institute of Plant Genome Research, Aruna Asaf Ali Marg, New Delhi-110067

# Present Address: National Agri-Food Biotechnology Institute, Mohali, Punjab

Running title: NAG plays a key role in galactinol synthase activity

*Corresponding author

To whom correspondence should be addressed: Dr. Manoj Majee
National Institute of Plant Genome Research, Aruna Asaf Ali Marg, New Delhi-110067, India; Email: manojmajee@nipgr.ac.in, Tel.91-11-26735193; Fax. 91-11-26741658

Keywords: Galactinol Synthase, Chickpea, active site, homology modelling, Molecular docking, raffinose family oligosaccharide (RFO), reactive oxygen species (ROS)

Abstract

Galactinol synthase (GolS) catalyzes the key regulatory step in the biosynthesis of Raffinose Family Oligosaccharides (RFOs). Even though the physiological role and regulation of this enzyme has been well studied, little is known about active site amino acids and the structure-function relationship with substrates of this enzyme. In the present study, we investigate the active site amino acid and structure function relationship for this enzyme. Using a combination of three-dimensional homology modelling, molecular docking along with a series of deletion, site directed mutagenesis followed by in vitro biochemical and in vivo functional analysis; we have studied active site amino acids and their interaction with the substrate of chickpea and Arabidopsis GolS enzyme. Our study reveals that the GolS protein possesses GT8 family-specific several conserved motifs in which NAG motif plays a crucial role in substrate binding and catalytic activity of this enzyme. Deletion of entire NAG motif or deletion or the substitution (with alanine) of any residues of this motif results in complete loss of catalytic activity in in vitro condition. Furthermore, disruption of NAG motif of CaGolS1 enzyme disrupts it’s in vivo cellular function in yeast as well as in planta. Together, our study offers a new insight into the active site amino acids and their substrate interaction for the catalytic activity of GolS enzyme. We demonstrate that NAG motif plays a vital role in substrate binding for the catalytic activity of galactinol synthase that affects overall RFO synthesis.
Introduction

Galactinol synthase aka inositol -3-α galactosyl transferase (GolS, EC: 2.4.1.123) is a member of Glycosyltransferase family 8 GT8 family glycosyltransferase in CAZy (Carbohydrate Active enzymes) database and is present only in flowering plants [1,2]. GolS catalyzes the first committed step in the biosynthesis of Raffinose Family Oligosaccharides (RFO) and is highly conserved among plant species. The enzyme essentially carries out the transfer of galactosyl moiety from uridine diphosphate galactose (UDP-alpha-D-galactose) to myo-inositol for the synthesis of galactinol (UDP + O-alpha-D-galactosyl-(1->3)-1D-myoinositol) which acts as a stepping stone for the biosynthesis of the RFO members [3]. This free galactinol essentially supplies activated galactosyl moiety to generate series of RFOs including raffinose, stachyose, verbascose by the enzymes raffinose synthase, stachyose synthase and verbascose synthase, respectively [4]. RFOs play a very important and diverse physiological roles in plants such as seed desiccation tolerance [5–9], translocation of photosynthetic assimilates [10], promote biotic [11,12] and abiotic stress tolerance [13–16]. Galactinol and RFOs are also believed to act as an osmolyte to maintain the cell turgor pressure and stabilizes the integrity of cellular proteins and membranes during stressful environments [17]. Most of the plants have more than one isoform which are usually encoded by a small gene family. In Arabidopsis, GolS isoforms are encoded by seven distinct genes, while in chickpea (Cicer arietinum) GolS are encoded by two genes [5,8,15]. GolS genes were shown to be differentially expressed in organs and in response to different stimuli to perform distinct physiological functions [4,5,14,18–20]. The overexpression of GolS in homo- or heterologous system confers abiotic stress tolerance in diverse species. For instance, the overexpression of AtGolS2 imparts drought stress tolerance in Arabidopsis as well as rice [15,21]. Further, we have previously established a direct role of chickpea GolS in abiotic stress tolerance and seed longevity by reducing the ROS mediated damage [5,18,22]. The GolS isoforms are apparently a monomeric protein with a calculated molecular weight of 35 kDa to 45 kDa. The GolS protein possesses GT8 family-specific motifs such as DxD, HxxGxxKPW and GXG which are predicted to be important for GolS enzyme activity, for instance DxD motif is crucial for divalent cation binding [2]. However, due to the lack of crystallographic and structural studies of GolS, little is known about the relationship between enzyme structure, active site amino acids and catalysis. Therefore, even though physiological role, regulation and basic biochemical properties of this enzyme have been elucidated in several plant species, studies pertaining to the identification of active site amino acid residues, structure-function relationship between substrates and enzymes are largely untapped.

In our previous study, we have reported five distinct GolS isoforms (CaGolS1, CaGolS1′ CaGolS2, CaGolS2’ and CaGolS2′′) in chickpea and CaGolS1′ (Accession no: KU189227), an alternative splice variant of CaGolS1, was found to be biochemically inactive [5]. These observations prompted us to investigate the biochemical and structural basis of the loss of catalytic activity of this CaGolS1′
isoform, and to investigate the potential functional domain/motifs and active site amino acid residues of GoLS protein essential for the catalytic activity.

In the present work, we describe the structural, bioinformatic and biochemical analysis of GoLS enzyme and explored the biochemical basis of the loss of catalytic activity of galactinol synthase. Further, in an attempt to identify specific motif/domain and/or amino acid sequence(s) crucial for the catalytic activity of GoLS enzyme, we performed in-depth bioinformatic analysis of CaGoS1 using several approaches followed by making a series of deletion or substitution mutants of the residues identified as potentially important for GoLS activity. Biochemical analysis of these mutant enzymes in *in vitro* demonstrated that NAG motif plays a very important role in the substrate binding and catalysis, where any point mutation in the NAG motif turn-out to be a total loss in GoLS enzyme activity. These findings were further confirmed by *in vivo* analysis in yeast and *Arabidopsis* and the findings were supported by *in silico* approach also.

**Experimental**

**Plant materials and growth conditions**

*Arabidopsis thaliana* ecotype Columbia (Col-0) was used in this study and were grown and maintained in a plant growth facility maintained at 22°C±2°C, RH of 60% with a 16h light (200µmol m⁻² s⁻¹ light intensity)/ 8h dark cycle [23].

**Molecular docking and insilico analysis**

Sequences of the CaGoS1 (*KU189226*) and AtGoS1 (*AT2G47180*) were retrieved from NCBI and TAIR respectively. Motif and domain searches were made on EBI server [http://www.ebi.ac.uk/Tools/InterProScan/](http://www.ebi.ac.uk/Tools/InterProScan/) employing InterProScan, and KEGG Motif Scan search server [https://www.genome.jp/tools/motif/](https://www.genome.jp/tools/motif/). Orthologs of other plants species were obtained from KEGG [http://www.genome.jp/kegg/](http://www.genome.jp/kegg/) and the same server was employed for multiple sequence alignments (CLUSTALW) [24], using the homologous plant proteins showing more than 80% sequence homology. Sequences of CaGoS1, CaGoS1’, and their respective deletion and substitution mutants were submitted for 3-dimensional structure predictions at I-TASSER server [http://zhang.bioinformatics.ku.edu/I-TASSER/](http://zhang.bioinformatics.ku.edu/I-TASSER/). Structures were selected on the basis of c-score values and in agreement with Ramachandran Plot using VMD software (University of Illinois). Molecular docking of 3D structures of selected proteins with UDP-galactose was performed using Patch Dock server [25] [http://bioinfo3d.cs.tau.ac.il/PatchDock/](http://bioinfo3d.cs.tau.ac.il/PatchDock/) and results were refined using FireDock server [26] [http://bioinfo3d.cs.tau.ac.il/FireDock/](http://bioinfo3d.cs.tau.ac.il/FireDock/). Interacting amino acid side chains, drug molecule orientation and docking feasibility was based on Fire Dock scores and visualizations with VMD software.
Construction of mutant variants

For generation of deletion variants, two strategies were employed i.e., SOE-PCR (Splicing by Overlap Extension-PCR) for larger deletion and DpnI based mutation for (1 aa InDel/mutations. The deletion variant of the CaGolS1 were generated by overlapping PCR in which two PCR were carried out i.e., primary and secondary PCR. In primary PCR, the two primer pairs (A/B and C/D) were used to amplify the CaGolS1 fragments, in which the primer B and C were having overlapping overhangs. The junction of these two overlapping primers (B and C) lack the nucleotide sequence to be deleted that facilitate the deletions (or the modified nucleotides that need to be incorporated) of the desired fragment. The resultant PCR product was purified, mixed in equimolar and used as template for secondary PCR amplification. The first-14-cycle of secondary PCR was carried out without primers, to accumulate the deletion product which was later amplified using full length primer pairs (A/D).

For incorporation of point mutations in CaGolS1, the CaGolS1 cloned in pET23b was used as template and inverse PCR was carried out. The desired mutations were incorporated in the primer set used for inverse PCR (oligonucleotide sequences are mentioned in Table S1). The PCR reaction was then subjected to DpnI digestion to remove the methylated DNA followed by purification and transformation in BL21DE3 as a protein expression host cell.

Bacterial overexpression and protein purification of CaGolS1 and its deletion variants

CaGolS1 and the deletion constructs were transformed in E. coli (Bl21DE3) cells were grown in LB medium supplemented with Ampicillin (100 μg ml⁻¹) at 37 °C with continuous shaking at 180 rpm. The recombinant protein was expressed by adding 0.5 mM (Isopropyl-β-d-thiogalactopyranoside) IPTG at 0.5 OD₆₀₀ and incubated at 6 h to 8 h at 30 °C with continuous shaking. The cells were then harvested at 12000 g for 3 min at 4 °C. Bacterial cells were lysed by freeze-thawing and sonication followed by purifying the protein fraction using Ni–NTA (nickel nitrilotriacetic acid) affinity chromatography as per the manufacture instructions (GE) and analyzed on SDS-PAGE [27–29].

GolS activity

Enzyme activity of galactinol synthase was measured by colorimetric method described previously [18][30]. The 100 μl assay mix [60 mM myo-inositol, 2 mM dithiothreitol (DTT), 50 mM HEPES buffer (pH 7.0), 4 mM MnCl₂, 20 μg of bovine serum albumin (BSA) and 4 mM UDP-galactose] were incubated for 60 mins at 32°C. The reaction was stopped by placing the tubes in boiling water bath for 2 min. To each tube 500 μL of water, 10 μL (0.3 U) of potato apyrase, and 150 μL of apyrase reaction mixture (Tris-HCl buffer, 250 mM [pH 7.5], 25 mM KCl, 7.5 mM CaCl₂, 0.5 mM EDTA-Na, and 50 mM glucose) were added and incubated at 37°C for 15 min. The apyrase reaction was terminated by addition of 60 μL of 75% TCA and immediately placed on ice for 15 min, followed by centrifugation at 10,000 g for 10 min. The amount of Pi released in the supernatant was determined...
by a modified Fiske and SubbaRow (1925) protocol [31]. To reaction mixture, 100 μL of 2.5% ammonium molybdate (dissolved in 2 N HCl) and 40 μL of Fiske and SubbaRow reducer were added. After 2 min incubation at room temperature, 40 μL of 34% sodium citrate.2H2O solution were added, and absorbance was immediately measured at 660 nm. The amount of Pi (inorganic phosphate) formed by the hydrolysis of UDP was determined using a standard curve constructed with KH2PO4 and correlated to the amount of UDP produced by the galactinol synthase.

Thermo-tolerance assay

For comparing the thermostolerance potential delivered by deletion variant of CaGoIS enzyme. We cloned the CaGoIS1, CaGoIS1Δ (N187A) in the pYES: DEST52 vector for yeast expression using gateway-based cloning (Invitrogen). The constructed pDEST52-CaGoIS1 and pDEST52 empty vector were separately transformed into yeast strain (INVSc1) using the PEG-lithium acetate-based transformation protocol. To inspect the CaGoIS activity of yeast cells transformed with pDEST52:CaGoIS1 and pDEST52:CaGoIS1Δ yeast cells were grown to logarithmic growth phase (OD=1). The yeast cell was centrifuged and washed, the cell pellet was lysed by vortexing with glass beads to prepare crude lysate [32]. The thermo- tolerance potential was also assessed using spot assay and growth curve of yeast cells harboring the pDEST52:CaGoIS1 and pDEST52:CaGoIS1Δ. For this, the yeast cells were grown to the mid-logarithmic phase of growing cells with A600 0.5 (about 1 x 10^7 cells ml^-1) on YEB medium at 30°C, and the cell density was adjusted to 0.02 (A600). For spot assay, the equal cell density (A600 0.02) was further serially diluted up to 10^-6 from each treatment and spotted (5 μl) on the plate in triplicate [33]. For control the plate was kept at 30°C while for thermal- stress treatment, the plate was kept at 42°C for 8 h then the plate was shifted to 30°C and the growth was monitored. Similarly, growth curve was plotted by assessing the growth (Absorbance OD600) of yeast cells incubate at 37°C and 30°C for every 2 hours. The initial cell density of all the yeast cells/treatment were kept similar i.e., OD = 0.2 A, and the growth was recorded at every 2 h by scoring the absorbance (OD600). Each treatment was set up for 3 repetitions, with at least four biological replicates.

Generation of transgenics

To generate transgenic lines of CaGoIS1 and CaGoIS1Δ (N187A), we cloned the CaGoIS1 in pENTR™/D-TOPO™ vector to generate entry clones. The entry clones were further subcloned in the destination vector (pEarlyGateway:201) using gateway cloning approach (Invitrogen). The constructs were confirmed through sequencing and transferred to agrobacterium cells (GV3101). Subsequently, floral dip was performed to generate the transgenic plants of CaGoIS1 and CaGoIS1Δ. The positive transformants were screened through Basta resistance (120 mg L^{-1}) till we obtained T3 homozygous
The T3 homozygous transgenic lines were further selected on the basis of increased mRNA level, GolS protein and galactinol content.

**Western blot analysis of CaGolS1 and CaGolS1Δ mutant lines**

Crude protein from transgenic (CaGolS1 and CaGolS1Δ) along with wild type and vector control were extracted using protein extraction buffer used in our previous study [35]. The concentration of isolated protein was determined using Bradford reagent (GE). For protein electrophoresis, 10 μg proteins were separated on 12% SDS-PAGE and then blotted onto PVDF (polyvinylidene difluoride) membrane using the Biorad blotting system. The membrane was incubated with the primary monoclonal Anti-HA antibody (1:2000 dilution) (Merck), followed by horseradish peroxidase-conjugated secondary antibody (GE Healthcare). As a loading control, tubulin was detected using an anti-tubulin antibody (Sigma) following the same procedure.

**RNA extraction and cDNA synthesis for Quantitative PCR assays**

Total RNA was extracted using the TRIzol reagent (Sigma) and the quantification of the RNA samples was assessed by using the ND-1000 UV-visible light spectrophotometer [36]. RNA samples with a 260/280-nm wavelength ratio of ~2 and a 260/230-nm wavelength ratio >2 was retained for analysis. Total RNA (1 μg) was reverse transcribed into cDNA, using the verso cDNA synthesis kit (thermo scientific). The absence of genomic DNA in RNA samples was checked by qPCR before cDNA synthesis. Also, a negative control (No Template Control) was incorporated in each assay. For normalization, the transcript levels of two endogenous reference genes (18S rRNA and EF1α) with stable expression were used in each assay. The gene specific primer set along with the reference genes used for the qRT-PCR were previously described and validated [37]

**Controlled Deterioration Test (CDT)**

The T3 transgenic seeds expressing CaGolS1 and CaGolS1Δ were subjected to a controlled deterioration test (CDT) to assess their vigor. In this, the seeds were initially imbibed in water to enhance their moisture content and later subjected to high temperature and relative humidity [5,35,38]. The treated seeds (4 days of CDT) were immediately used for germination assay and viability test (TTZ assay). While for MDA and H2O2 assay, the seeds were used immediately or snap freeze in liquid nitrogen and stored in -80°C for later use.

**Germination and TTZ assay**

The germination and tetrazolium assay (TTZ; 2,3,5-Triphenyl-tetrazolium chloride) were essentially performed as described previously [5]. After CDT treatment, the seeds were evaluated for their germination potential. For this, the seeds were placed on ½ MS medium and kept in growth chamber. The seeds were monitored for germination potential. The seeds were considered germinated and
scored after the 1mm radicle emersions. Besides, the seeds were subjected to 1% TTZ stain to
determine the viability of seeds before and after CDT and photographed [23,39,40].

**Malondialdehyde and H₂O₂ estimation**

T3 seeds of transgenic along with WT and VC, challenged with CDT was assayed for lipid
peroxidation based on TBA (thio-barbituric acid) method and H₂O₂ using potassium iodide (KI)
method as essentially performed previously [14,27].

**GC-FID**

To determine the galactinol, raffinose and myo-inositol content in transgenic lines, we used GC-FID
approach as previously described [14,41]. Briefly, polar metabolites were extracted, lyophilized and
derivatized for silylation using MSTFA. The GC-analysis was conducted with Shimadzu GC-2010
system occupied with FID [14].

**Statistical analysis**

The data presented in the manuscript represent the mean value and standard deviation of triplicate
analysis. The one-way analysis of variance (ANOVA) was employed to compare the statistically
significant difference between the sample and control value using DMRT as the post hoc test. The
difference in the mean was considered statistically significant if the $\alpha < 0.05$.

**Results**

**Structural and bioinformatic analysis of CaGolS1 predicts probable active site(s)/domain(s)/motif(s)**

As, the 3D structure for galactinol synthase is not available, we generated the same through I-tasser
server as described in Experimental procedure section. The best model structure was selected as per
the I-tasser confidence level (c-score), in agreement with Ramachandran plot, and was used for
further bioinformatics analysis (Table 1). The structural analysis of CaGolS1 typically showed a
complex structure with seven $\alpha$-helices and nine $\beta$-sheets forming the central core area of protein
surrounded by several varied length polypeptide chains on surface of protein. The central core of
protein made of $\beta$-sheets surrounded by $\alpha$-helices forms proper binding cavity in the central region of
protein for any suitable substrate (Fig. 1 A-B). Several conserved domains/motifs such as FLAG,
DxD, HxxGxxKPW, APSAA and NAG which are reported previously [2,5] are also shown in Fig.
1C. In order to further identify the presence of any other conserved sequences in CaGolS1, we
performed the InterProScan and KEGG Motif Finder search by submitting the complete sequence to
both servers separately. InterProScan showed almost complete sequence conserved for GT8 family
with residues from 29-271 (PFAM: PF01501) whereas PANTHER (PTHR11183:SF89) shows entire
sequence conserved for Galactinol synthase, Gene3D shows sequence from residue 22 – 300
conserved for Nucleotide-diphospho-sugar transferases (G3DSA: 3.90.550.10) and SUPERFAMILY
shows domain conserved for Nucleotide-diphospho-sugar transferases with residues from 28 – 299
(SSF:53448). Meanwhile, in a parallel search by KEGG Motif Finder, it shows sequence with
residues from 29 – 271 (PF01501) conserved for GT 8 and one more domain conserved for Mannosyl
transferase putative (Mannosyl_trans3) with residues from 36 – 138 (PFAM: PF11051). Structural
analysis by I-tasser, predicted good binding with UDP in the central cavity (Table 2) of the protein
with interacting residues ranging from almost all above reported conserved motifs/domains, and this
data also corresponds to the residues predicted by InterProScan as substrate binding sites. Thus, it is
clear that CaGolS1 interacts properly with UDP galactose and residues from reported motifs/domains
are either completely or partially involved in the interaction. Nevertheless, I-tasser and InterProScan
both have identified some more residues other than the region of reported motif/domain as potential
substrate binding site.

Comparative analysis of catalytically active CaGolS1 and inactive CaGolS1’ reveals loss of
probable active site and conformational integrity of CaGolS1’

In our previous study, we have reported that CaGolS1’, a splice variant of CaGolS1, is enzymatically
inactive, and the lack of biochemical activity of CaGolS1’ is conceivably due to the absence of 73aa
stretch (Fig. S1) [5]. This gives us an insight that residue(s) from 73aa stretch are likely to be
responsible to impart GolS activity to the molecule. To reconfirm that the deletion of these 73aa
indeed caused CaGolS1’ isoform to be biochemically inactive, we generated a deletion construct
where this 73aa stretch was deleted from the biochemically active wild type CaGolS1 (designated as
CaGolS1Δ1). Subsequently, CaGolS1Δ1 was bacterially over expressed with C-terminal hexa-his-tag
and recombinant proteins were purified to homogeneity using pre-packed Ni-NTA column. Purified
recombinant protein was assessed for GolS activity and results revealed that CaGolS1Δ1 was indeed
biochemically inactive (Fig. S2). Next, to investigate whether the deletion of this stretch of 73aa
residues had changed the structural conformation and/or disrupted the active site amino acid residues,
we generated the 3D structure of CaGolS1’ using I-tasser server as described in Experimental
procedures section. It is important to keep into consideration that like other biochemically active
GolS, CaGolS1 and CaGolS1’ (Accession no: KU1892277 both possess FLAG, DxD, HxxGxxKPW
and APSAA conserved motifs and sequences (Fig. 1E). 3D model structure of CaGolS1’ which lacks
a stretch of 73aa showed five α-helices and seven β-sheets where five pairs of β-sheets were
surrounded by α-helices and several intermediary placed varied length polypeptide chains invariably
present on surface of the protein (Fig. 1D). Data obtained by using ligand-binding site prediction by
InterProScan and I-tasser showed a significant decrease in number of amino acids involved in
probable active binding site formation for any suitable substrate in this mutant variant. Furthermore, data obtained from I-tasser and InterProScan showed that almost 30-35% of residues predicted as probable substrate-binding site belongs to the region of 73aa. This structural analysis between these two proteins established that wild type CaGolS1 is much more stable and active protein with a greater possibility to interact and bind with substrate UDP-Galactose/Inositol as compared to its splice variant. Furthermore, 3-dimensional structural and their docking analysis with GolS substrate viz UDP-galactose revealed that CaGolS1 wild type protein showed proper binding of UDP-galactose in the central core cavity of the protein, whereas this substrate interacts poorly with surface residues of CaGolS1' protein (Fig. 2 A-B). Molecular docking analysis of CaGolS1 with UDP-galactose reveals that nearly 50% of interacting residues belong to the region of 73aa. Therefore, it is evident that these markedly displaced loops and folds in structure of splice variant (CaGolS1') serves as deterrent against the substrate binding and results in an apparent conformational change in the active site which might be attributed to the loss of catalytic activity of this isoform. Various parameters used for bioinformatic analysis and molecular docking are represented in Table 1.

Homology modelling and deletion study reveal key motifs for GolS activity.

Based on previous analysis, it is assumed that 73aa regions possess important motifs/residues which are likely to play important role in substrate binding for the catalysis. To identify the precise potential key domain(s)/motif(s) important for catalysis or imparting GolS activity, we next performed in-depth bioinformatic analysis. Initially, BLASTP analysis was carried out using NCBI server and protein was found to be conserved as GolS1 in several plant species. Next, we carried out the Multiple Sequence Alignment (MSA) analysis with plant species showing more than 80% sequence homology to CaGolS1, all their sequences we retrieved and submitted to ClustalW for MSA which showed the presence of all reported conserved domains and motifs in all selected species (Fig. S3). Thereafter, we searched for the conserved sequences/residues from the stretch of 73aa region and identified LYFNAG (184, 185, 186, 187, 188, 189), FAEQDF (215, 216, 217, 218, 219, 220) and YNLVLAMLW (236, 237, 238, 239, 240, 241, 242, 243, 244) as conserved regions in all the selected species (Fig. S3). Subsequently, to identify which of these conserved domains are important for GolS activity, we generated four deletion variants of CaGolS1 (CaGolS1Δ2, CaGolS1Δ3, CaGolS1Δ4, CaGolS1Δ5) by deleting 18-19aa sequence spanning these conserved sequences from 73aa patch (Fig. S4) using splicing by overlap extension PCR (SOE-PCR). In these 18-19aa deletion mutants, LYFNAG belongs to the first deletion mutant, FAEQDF belong to third deletion mutant and YNLVLAMLW is present in fourth deletion mutant whereas no conserved sequences were observed in second deletion mutant. All these deletion mutants were bacterially expressed and then purified enzymes were used for GolS assay. Our results showed that all deletion mutants were enzymatically inactive (Fig. 3A-C). It can be predicted that these conserved sequences, incomplete or in partial, are playing an important role in
imparting the catalytic activity of the enzyme. Though the loss of enzyme activity of all four variants can also be due to the altered conformation occurred by the larger deletion of 18-19aa. Thus, the deletion of relatively larger patch of 18-19aa (CaGolS1Δ2, CaGolS1Δ3, CaGolS1Δ4, CaGolS1Δ5) essentially leads to a conformational change and was not confirmatory to identify the important motif and amino acids that are crucial for the catalytic activity of GolS enzyme. Hence, to narrow down to identify precise amino acid residues/motifs responsible putative substrate-binding site from these identified conserved domains (LYFNAG, FAEQDF, YNLVLAMLW), we cross-checked for any probable ligand-binding residues predicted/identified by I-tasser or InterProScan, and it was observed that NAG and FAE have been predicted by both I-tasser and InterProScan as probable substrate-binding site. More importantly on analyzing the same through selected 3D model structure we identified that particularly NAG of LYFNAG conserved motif in the central core actively take part in forming the substrate-binding cavity of GolS.

**Mutation in NAG motif causes a complete loss of GolS activity**

To verify such possibilities, we made bacterially expressed recombinant mutant proteins with these deletions [CaGolS1Δ6: NAG and CaGolS1Δ7: FAE] (Fig. 4A-B). Recombinant proteins were purified to homogeneity using pre-packed column of Ni-NTA matrix as described in previous section, and subsequently analyzed their activity. Among these mutant variants, CaGolS1Δ6 (NAG) showed the complete loss of activity whereas CaGolS1Δ7 (FAE) exhibited nearly 50% reduction in catalytic activity compared to CaGolS1 (Fig. 4C). These results give a strong indication that NAG is crucial for GolS activity whereas the FAE might be indirectly responsible for activity as in FAE mutant NAG will be at its native position but due to the deletion of FAE the conformation of molecule is slightly deteriorated, making changes in the active site. To revalidate these results, we performed the molecular docking of both mutants with UDP galactose and found that CaGolS1Δ6 interacts poorly with substrate on the surface of the molecule indicating that deletion of NAG results in the deformation of active site leaving substrate no proper binding cavity whereas CaGolS1Δ7 shows below moderate level of interaction with UDP suggesting that deletion of FAE resulted in conformational change blocking the path of UDP to reach the binding cavity/site of the molecule, and thus resulted in the reduction of GolS activity by ~50% (Fig. 2 C-D). Moreover, on looking into the details of molecular docking analysis of wild type CaGolS1 with UDP galactose it was found that NAG residues showed proper binding with substrate (Fig. 2a). Thus, it can be predicted that NAG (N187, A188, G189) is crucial for imparting GolS activity to the molecule.

To further investigate the catalytic importance of each amino acid residue in NAG motif (N187 A188 G189) in detail, deletion and replacement of these residues have been carried out using SDM approach. In N187A [CaGolS1Δ8] or G189A [CaGolS1Δ9] mutants GolS activity was completely abolished (Fig. 4c). Further, we deleted either N187(CaGolS1Δ10), A188 (CaGolS1Δ11) or G189 (CaGolS1Δ12)
and assessed the effect of each residue on enzyme activity (Fig. 4c). Our results clearly depicted that all the amino acids of NAG motif are critically important for GolS activity.

The NAG motif is highly conserved across the plant species, therefore, to confirm whether NAG motif is also important for GolS activity in other plant species, we performed similar approach of SDM and disrupt the NAG motif and generated constructs for three different mutant variants of AtGolS1 (AtGolSΔ1: NAG to AAA; AtGolSΔ2: N193A; AtGolSΔ3: N193 deleted). Arabidopsis GolS1 (AtGolS1) along with its three mutant variants were bacterially expressed and purified using affinity chromatography, and subsequently catalytic activity was examined and compared with wildtype AtGolS1 (Fig. 5a-c). As anticipated, all AtGolSΔ1; AtGolSΔ2 and AtGolSΔ3 mutant variants were appeared to be enzymatically inactive, signifying the vital importance of intact NAG motif for GolS activity across the species.

**Effect of N187 mutation of CaGolS1 function in yeast**

To lend stronger support for the above finding, we further examined our results in vivo. Previously we reported that CaGolS1 improves thermal stress tolerance to *Arabidopsis* by limiting the heat stress-induced ROS level [14]. Considering the role of CaGolS1 in heat stress tolerance, we assessed the growth pattern of yeast (INVSc1) harbouring native CaGolS1, CaGolS1Δ (N187A) and empty vector control under thermal stress as mentioned in material and method section. Initially, we conducted the western blot analysis to confirm that CaGolS1 and CaGolS1Δ (N187A) both are indeed expressed in the yeast cells (Fig. S5). At 30°C, yeast cells transformed with CaGolS1, CaGolS1Δ (N187A) and empty vector grew equally well and did not show a significant difference in their growth pattern (data not shown). However, yeast cells expressing CaGolS1 exhibited better growth response than empty vector control yeast cells after heat stress exposure. After heat stress treatment, the growth of empty vector transformed cells was inhibited, while the CaGolS1 transformed cells were noticeably less inhibited in similar condition and able to mitigate heat stress induced growth inhibition (Fig. 6).

Similarly, spot assay after thermal stress also revealed that yeast cells transformed with mutant construct CaGolS1Δ (N187A) were not able to mitigate heat stress induced growth inhibition and grew rather similar to empty vector transformed yeast cells. Even though the deletion variant of CaGolS1 protein i.e. N187A was expressed in yeast, the mutant CaGolS1 transformed cells were sensitive to heat stress like empty vector control cells. These results suggest that yeast cells expressing biochemically active GolS were able to exhibit heat stress tolerance but not those cells which express biochemically inactive GolS. For further confirmation, we compared the GolS activity among CaGolS1, CaGolS1Δ- and vector-transformed yeast cells and observed that GolS activity was only present in pDEST52:CaGolS1 transformed yeast, while pDEST52:CaGolS1Δ and pDEST52: empty vector showed similar pattern. This data strongly suggests that point mutation in NAG motif results
in biochemically inactive GolS which is unable to provide thermotolerance to yeast in contrast to its
native GolS (Fig. 6).

Disruption of N187 residues in NAG motif in CaGolS1 significantly affects its function on seed
vigor and longevity in planta

Previously, Arabidopsis transgenic plants constitutively overexpressing CaGolS1 were shown to
accumulate significantly increased galactinol and raffinose which consequently improves seed vigor
and longevity [5]. In our previous sections, we evidently demonstrated that a point mutation in NAG
motif of GolS1 results in complete abolition of enzymatic activity in vitro and in vivo. Here, we
intended to further establish that NAG motif is indeed critical for galactinol synthase activity for the
synthesis of galactinol in planta. To address this, we chose CaGolS1Δ (N187A) mutant for further in
planta analysis as our in vitro results have confirmed that any point mutation in NAG motif causes
loss of GolS activity. Arabidopsis transgenic lines overexpressing the N187A mutant as well as its
native biochemically active CaGolS1 with HA tag was generated under the control of constitutive
promoter (CaMv35S). Initially, overexpression of CaGolS1 and CaGolS1Δ (N187A) was confirmed
through transcript accumulation (Fig. S6) of the respective transgenic lines followed by the GolS
activity in three independent overexpression lines (OE-1, OE-2, OE-3) (Fig. 7A). Further, western
blot analysis was carried out to confirm that CaGolS1 and CaGolS1Δ (N187A) were indeed expressed
in respective transgenic lines (Fig. S7). In biochemical and metabolic analysis of the transgenic lines,
we found a significantly increased total GolS activity and accumulation of galactinol as well as
raffinose were observed in CaGolS1 transformed lines compared to wild type or vector transformed
lines (Fig. 7, Fig. S8). However, in contrast to CaGolS1 transformed plants, transgenic plants
transformed with CaGolS1Δ did not show any increase in GolS activity, galactinol and raffinose
content despite the protein was expressed. The galactinol and raffinose content of CaGolS1 mutant
transformed line were found fairly similar to wild type or empty vector control lines. Subsequently, a
Controlled Deterioration Test (CDT) was conducted to evaluate the seed vigor and longevity for these
transgenic lines. As anticipated, seeds from CaGolS1 transformed lines exhibit improved seed vigor,
whereas the CaGolS1- N187A transformed seeds behaved similar to the wildtype counterpart. After 4
days of CDT, CaGolS1 transformed seeds exhibited 50% to 60% germination while CaGolS1Δ
transformed seeds showed only 10-12 % germination like control seeds (wild type and vector control)
(Fig. 7 B-D). After CDT, only CaGolS1 transformed seeds showed dark red staining in contrast to
CaGolS1Δ (N187A) transformed seeds and control seeds, which remained unstained or stained pale
red (Fig. 7E). These results clearly indicated that only CaGolS1 transformed seeds were viable after
aging treatment due to overexpression of catalytically active CaGolS. As in our previous study, we
found a positive influence of galactinol and raffinose on ROS scavenging during seed aging [5], it is
tempting to speculate that whether the reduced seed germination of CaGolS1Δ overexpressing seeds is
accompanied by an overaccumulation of ROS level. To address this point, we determined the $\text{H}_2\text{O}_2$
and MDA content in the transgenic and control seeds after CDT exposure (Fig. 7 F, G). In the WT
and CaGolS1Δ mutant seeds, both $\text{H}_2\text{O}_2$ and MDA have been found to be significantly higher in
response to CDT. In general, the mutant variant and control seeds behaved almost similar, also the
reduced germination in CaGolS1Δ mutant were attributed to higher ROS accumulation which could
indicative of higher seed deterioration after aging. These data clearly suggested that N187A mutant
generate an enzymatically inactive galactinol synthase that could not contribute to the accumulation of
galactinol and its subsequent product and thereby provide a straightforward explanation for its
observed phenotype of reduced seed vigor. Thus, its overexpression did not contribute to seed vigor
and longevity and showed reduced germination which is strongly associated with increased cellular
death after CDT resulting from the reduced ROS scavenging capacity.

**Discussion**

Based on their sequence homology, glycosyl transferase (GT) have been categorized in hundreds of
different classes in the CAZy (carbohydrate active enzyme) database[42]
(http://www.cazy.org/GlycosylTransferases.html). GT catalyzes the synthesis of glycoconjugates by
transferring the activated sugar moiety from the donor molecule to its acceptor [43,44]. In the
evolutionary spectrum, GolS is restricted to the plant kingdom and belongs to GT8 family [4,16]. The
GolS enzyme catalyzes the key regulatory step of RFO biosynthesis and has led to raise substantial
interest in understanding the RFO regulation in plants. Even though the physiological role and
regulation of this enzyme have been characterized from several plant species, studies on determining
active sites and structure-function relationship have not been carried out so far [45–49]. To date, no
crystallographic studies are available which shed light on the active site and catalytic mechanisms. In
this work, we have identified that NAG motif is critically important for GolS catalytic activity
through homology modelling and site directed mutagenesis followed by *in vitro* and *in vivo* functional
characterization. Homology modelling has now been regularly used to understand the sequence-
structural relationship. Sequence analysis reveals that GolS proteins possess DxD, HxxGxxKPW
motifs and conserved sequences like NAG, APSAA. Among these, NAG, DXD amino acid residues
are predicted to be a part of GolS enzyme active site and further DxD motif is also shown to be
required for divalent cation binding [2]. In our previous studies [5,14], we have identified the
CaGolSI’ isoform possess DxD, HxxGxxKPW motifs but lacks NAG sequence. CaGolS2’ (Accession
no: KU214572) is 678 bp in length and encodes 228aa protein while CaGolS2” (Accession no:
KU214573) is 657 bp in length and encodes 218aa protein. Both these isoforms contain HxxGxxKPW
motif but lack NAG and DxD motif.
In this study, we have focused on the biochemical aspect of CaGolS and pinpoint the crucial residues involved in the catalytic activity of CaGolS. To address this, we generated several deletion variants using SDM approach of several residues to assess the functional consequence at their enzyme activity. When we inspected the CaGolS1 and CaGolS1’ at structural level, there were several structural deformities present in the CaGolS1’ which might be correlated with the lost activity of CaGolS1’. The 3D structure of CaGolS1 shows the complex structure formed of seven α-helices and nine β-sheets and varied length polypeptide chains in coil and/or linear form. The ligand-binding residues predicted by InterProScan and I-tasser lies in the core of the protein molecule making a suitable binding cavity/site for substrate, whereas in CaGolS1’ structure is formed of five α-helices and seven β-sheets and intermediary placed varied length polypeptide chains invariably present on the surface of the protein. Also, a significant reduction in the number of probable ligand-binding residues was observed by the data retrieved from InterProScan and I-tasser which is believed to be due to the deletion/absence of 73aa stretch as compared to wild type molecule and is expected to be responsible for the deformities in the conformation of CaGolS1’ making it less suitable for interaction with any suitable substrate. To get a further in-depth view of the structural conformation we performed the comparative structural alignment analysis of both molecules by aligning both structures parallel to each other, but as both structures are complex and any deviation in conformation was difficult to identify so, we performed a step by step comparative structural alignment analysis by visualizing only first hundred aa of both molecules parallelly color-coded by different colors and hiding the rest of the structure of both molecules in VMD followed by next hundred aa (101 - 200) and hiding rest of structure and finally visualizing aa from 201 to last aa of both molecules and hiding rest of structure (Fig. S9). In this analysis, we observed that structural conformation of molecules formed by first hundred aa is almost comparable with least deviation, but in the next two segments of analysis substantial deviations/changes were observed in CaGolS1’ as compared to CaGolS1. This is believed to be due to the absence/deletion of 73aa stretch which results in the loss of GolS activity in CaGolS1’ and also on performing the molecular docking with UDP galactose, UDP interacts poorly with CaGolS1’ as compared to CaGolS1 (Fig. 2).

Besides, the disruption of NAG domain in AtGolS enzyme evidently support our results found in CaGolS, and indicate that the NAG domain is indeed crucial for GolS among different species. The higher thermotolerance in the yeast cells expressing the CaGolS1 supported the fact that, despite being a plant-specific enzyme GolS play a significant role in thermal stress in yeast. In contrast, the yeast cells expressing the N187A mutant which is proved to be enzymatically inactive, behaved similar to the cell harbouring the empty vector (Fig. 6). We can correlate that the thermotolerance capability in yeast is indeed implicated through the presence of GolS. Further, the in vivo function assessment of NAG domain in planta also evidently supported the important role of NAG motif. As the CaGolS1Δ (disrupted NAG motif by N187A mutation) doesn’t impart the improved seed vigor as
that of CaGolS1 after accelerated aging. The deferential behaviour of the N187A-OE and CaGolS1-OE seeds in response to control deterioration test can be well correlated with the ROS accumulation in the seeds after CDT. The CaGolS1-OE accumulated higher galactinol/raffinose and lesser the ROS while the CaGolS1N187A-OE behaves similar to that of the wild type seed. The lower galactinol/raffinose and higher ROS level can explain the higher deterioration of WT and CaGolS1N187A-OE seeds. Together, results of our \textit{in vitro} and \textit{in vivo} analysis demonstrated that NAG motif plays a crucial role in substrate binding and enzyme catalysis. Overall, our results offer a new insight into the active site amino acids and structure function relationship of戈尔S enzyme where NAG motif plays a vital role in substrate binding for the catalytic activity of galactinol synthase.
Data availability

Gene and protein sequences are available in NCBI Genbank [CaGolS1: Accession no: KU189226 and CaGolS1:Accession no: KU189227]. All data and materials can be available upon request to corresponding Author (Manoj Majee).

Supporting Information: This work contains supporting information.

Acknowledgments

This work was supported by the NGCP (grant BT/AGR/CG-Phase-II/01/2014) and the core grant of National Institute of Plant Genome Research, Department of Biotechnology (DBT), and Government of India. AH thank University Grant Commission, Government of India, for research fellowships. The authors are thankful to DBT-eLibrary Consortium (DeLCON) for providing access to e-resources

Author contributions

MM coordinated the project, conceived and edited the manuscript. PS designed and conducted the major experiments, generated and analyzed the data, and wrote the manuscript. NUK helped in the western blot and conducted few biochemical analyses of deletion variants. AH helps in bacterial expression and purification. BK conducted the molecular docking and in silico analysis. All authors have read and approved the final manuscript

Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article.
References


Figure Legend

Figure 1. 3D model structure of (A) CaGolS1 (secondary structure mode), (B) CaGolS1 (Surf mode) with zoomed image of binding cavity in square box, (D) CaGolS1' (secondary structure), (C and E) CaGolS1 and CaGolS1’ showing the presence of reported conserved domains respectively. Pink colored residues (surf mode) – FLAG domain, yellow colored residues (surf mode) – DxD domain, green colored residues (surf mode) – HxxGxxKPW domain and red colored residues (surf mode) – APSAA domain, cyan colored residues (surf mode) – NAG domain

Figure 2. Molecular docking of UDP-galactose with 3D model structures of (A) CaGolS1, (B) CaGolS1’, (C) CaGolS1-NAG deleted mutant and (D) CaGolS1-FAE deleted mutant. Crystal structure represents UDP galactose, orange color shows interacting residues and green color shows NAG residues

Figure 3. Bacterial expression, purification and enzymatic analysis of CaGolS1 deletion mutants.
(A) SDS PAGE analysis of deletion variants of CaGolS1 expressed protein in E. coli BL21 (DE3), (B) represents the respective purified fraction of deletion variants of CaGolS1 using affinity chromatography [Empty vector - pET23b empty vector transformed induced cells; M- molecular weight marker; P- Pellet fractions; S- Soluble fractions]. (C) Comparison of enzyme activity among CaGolS1, CaGolS1Δ2, CaGolS1Δ3, CaGolS1Δ4, CaGolS1Δ5. Five μg of purified protein was used for GolS assay. Significant differences among means (α = 0.01) are denoted by the different letters.

Figure 4. In vitro analysis of the importance of NAG motif for GolS activity of CaGolS1 enzyme
(A) SDS PAGE analysis of deletion variants of CaGolS1 expressed protein in E. coli BL21 (DE3), (B) represents the respective purified fraction of deletion variants of CaGolS1 using affinity chromatography [Empty vector - pET23b empty vector transformed induced cells; M- molecular weight marker; P- Pellet fractions; S- Soluble fractions]. (C) Comparison of enzyme activity among CaGolS1, CaGolS1Δ6, CaGolS1Δ7, CaGolS1Δ8, CaGolS1Δ9, CaGolS1Δ10, CaGolS1Δ11, CaGolS1Δ12. Five μg of purified protein was used for GolS assay. Significant differences among means (α = 0.01) are denoted by the different letters. All the data points used for the bar graph are shown with circles.

Figure 5. In vitro analysis of the importance of NAG motif for GolS activity of AtGolS1 enzyme
(A) SDS-PAGE analysis of AtGolS1 and its deletion variants recombinant proteins expressed protein in E. coli BL21 (DE3). (B) represents the purified fraction of AtGolS1, AtGolSΔ1, AtGolSΔ2 and AtGolSΔ3 [Empty vector - pET23b empty vector transformed induced cells; M- molecular weight marker; P- Pellet fractions; S- Soluble fractions]. (C) Comparison of enzyme activity among AtGolS1, AtGolSΔ1, AtGolSΔ2 and AtGolSΔ3. Five μg of purified protein was used for GolS assay. Significant
differences among means ($\alpha=0.01$) are denoted by the different letters. All the data points used for the bar graph are shown with circles.

**Figure 6.** Effect of N187A mutation of CaGolS1 function in yeast.

(A) Comparative GolS activity from crude extract of yeast cells harbouring the CaGolS1, CaGolS1Δ (N187A) and empty vector grown under control and heat stress treatment. (B) The growth curve of yeast cells harbouring different construct (CaGolS1, CaGolS1Δ (N187A) or empty vector) during thermal stress exposure. (C) Spot assay represent the comparative growth of respective yeast cells subjected to thermal stress. Five microliters (5 µl) of yeast sample was used for spotting in tenfold serial dilutions for each case. All the data points used for the bar graph are shown with circles.

**Figure 7.** Functional assessment of N187A mutant [CaGolS1Δ] in plant.

(A) GolS activity was determined in three independent transformed lines of CaGolS1 [CaGolS1-OE1, CaGolS1-OE2, and CaGolS1-OE3], mutant CaGolS1Δ- [CaGolS1Δ-OE1, CaGolS1Δ-OE2 and CaGolS1Δ-OE3], wild type (WT) and vector control (VC). Fifty µg of crude protein was used for the assay. (B) Germination assay of transgenics seeds under control condition and (C) seeds subjected to CDT (45 °C and 100% RH) for 4 days. (D) Percent germination of CaGolS1, CaGolS1Δ, WT and VC transformed *Arabidopsis* seeds before and after CDT. Data are means ± SD of three biological repetitions. (E) Seed viability was analysed before and after CDT treatment using 1% TTZ staining. Quantitative analysis of (F) H$_2$O$_2$ and (G) MDA content in the respective seeds of transformed lines before and after CDT treatment. Data are means ± SD of three biological repeats. Significant differences among means are denoted by the different letters. All the data points used for the bar graph are shown with circles.

**Supplementary Information**

**Figure S1.** Sequence comparison of CaGolS1 and CaGolS1’ showing the deletion segment in the splice variant.

**Figure S2.** (A) SDS PAGE analysis of CaGolS1, CaGolS1’ and CaGolS1Δ1 recombinant expressed protein in *E. coli* BL21 and (B) Comparative enzyme activity of CaGolS1, CaGolS1 ‘and CaGolS1Δ1. 5µg of purified protein was used for GolS assay. Significant differences among means ($\alpha = 0.01$) are denoted by the different letters.

**Figure S3.** Multiple sequence alignment of galactinol synthase protein sequence from different species highlighting the conserved domain.

**Figure S4.** Highlighting the deletion segments of CaGolS1Δ1, CaGolS1Δ2, CaGolS1Δ3, CaGolS1Δ4, and CaGolS1Δ5 by deleting 18-19aa sequence from 73aa patch.
**Figure S5.** Western blot analysis of protein extracts from yeast cell transformed with CaGolS1 or CaGolS1Δ (N187A).

**Figure S6.** Quantitative RT PCR analysis of CaGolS1 and CaGolS1Δ overexpressing transgenic lines.

**Figure S7.** Western blot analysis of CaGolS1 and CaGolS1Δ overexpressing transgenic lines.

**Figure S8.** Quantification of galactinol and raffinose content in the transgenic lines overexpressing CaGolS1 and CaGolS1Δ.

**Figure S9.** 3D model structure of (A) CaGolS1 (yellow color), and (B) CaGolS1’ (cyan color); (C) comparative sequence alignment of 1-100aa of CaGolS1 and CaGolS1’, (D) comparative sequence alignment of 101-200aa of CaGolS1 and CaGolS1’, (E) comparative sequence alignment of 201-last aa of CaGolS1 and CaGolS1’

**Table S1.** Primers used in this study.
Table 1: 3D modelling and docking parameters used for bioinformatic analysis

<table>
<thead>
<tr>
<th></th>
<th>C-Score</th>
<th>TM-Score</th>
<th>RMSD Value (Å)</th>
<th>Global Energy</th>
<th>Attractive van der Waals forces</th>
<th>Repulsive van der Waals forces</th>
<th>ACE</th>
<th>Interacting residues with UDP</th>
</tr>
</thead>
<tbody>
<tr>
<td>CaGolS1</td>
<td>-1.06</td>
<td>0.58±0.14</td>
<td>8.8±4.6Å</td>
<td>-33.24</td>
<td>-20.74</td>
<td>4.40</td>
<td>-3.63</td>
<td>102, 106, 122, 124, 147, 148, 174, 176, 177, 178, 179, 187, 188, 189, 216, 217, 218, 239, 259, 261, 262, 322</td>
</tr>
<tr>
<td>CaGolS1'</td>
<td>-1.34</td>
<td>0.55±0.15</td>
<td>8.9±4.6Å</td>
<td>-32.63</td>
<td>-22.22</td>
<td>10.51</td>
<td>-5.06</td>
<td>7, 9, 12, 16, 17, 21, 49, 56, 78</td>
</tr>
<tr>
<td>NAG deleted mutant</td>
<td>-1.42</td>
<td>0.54±0.15</td>
<td>9.7±4.6Å</td>
<td>-39.73</td>
<td>-27.24</td>
<td>5.36</td>
<td>-2.20</td>
<td>37, 124, 164, 170, 172, 173, 186, 211, 213, 215, 256, 258, 259, 260, 261, 263, 322, 328, 329</td>
</tr>
<tr>
<td>FAE deleted mutant</td>
<td>-1.18</td>
<td>0.57±0.15</td>
<td>9.1±4.6 Å</td>
<td>-46.01</td>
<td>-19.80</td>
<td>3.03</td>
<td>-11.83</td>
<td>37, 99, 102, 168, 172, 212, 213, 214, 260, 259, 263, 327, 330, 331</td>
</tr>
</tbody>
</table>

Table 2: Substrate binding site prediction by I-tasser and InterProScan. Residues from 73aa region are shown in bold

<table>
<thead>
<tr>
<th></th>
<th>I-tasser</th>
<th>InterProScan</th>
</tr>
</thead>
<tbody>
<tr>
<td>CaGolS1</td>
<td>30, 31, 33, 37, 102, 103, 107, 122, 123, 124, 187, 188, 189, 214, 216, 217, 218, 259, 261, 262, 266</td>
<td>102, 103, 106, 120, 122, 123, 124, 147, 187, 188, 189, 217, 218, 240, 241, 259, 261, 263, 266</td>
</tr>
<tr>
<td>CaGolS1'</td>
<td>30, 31, 33, 34, 37, 122, 123, 124, 146, 150, 151, 152, 186, 188, 189, 193</td>
<td>102, 103, 106, 120, 122, 123, 124, 147, 186, 188, 189, 193</td>
</tr>
</tbody>
</table>
Figure 3

Activity (nmoles mg⁻¹ min⁻¹)

- CaGolS1
- CaGolS1Δ2
- CaGolS1Δ3
- CaGolS1Δ4
- CaGolS1Δ5

Empty Vector
Figure 4

Activity (nmoles mg⁻¹ min⁻¹)

- CaGolS1
- CaGolS1Δ6
- CaGolS1Δ7
- CaGolS1Δ8
- CaGolS1Δ9
- CaGolS1Δ10
- CaGolS1Δ11
- CaGolS1Δ12

Empty Vector

CaGolS1
CaGolS1Δ6
CaGolS1Δ7
CaGolS1Δ8
CaGolS1Δ9
CaGolS1Δ10
CaGolS1Δ11
CaGolS1Δ12
Figure 5

(A) SDS-PAGE gel showing the protein bands for different samples:
- Empty Vector
- AtGoS1
- AtGoS1.1
- AtGoS1.2
- AtGoS1.3

(B) Enlarged view of the gel bands for each sample:
- M (Marker)
- AtGoS1
- AtGoS1.1
- AtGoS1.2
- AtGoS1.3

(C) Activity assay results:
- Activity in umoles mg\(^{-1}\) min\(^{-1}\)

- AtGoS1
- AtGoS1.1
- AtGoS1.2
- AtGoS1.3
Figure 7