International Journal of Bioinformatics Research
ISSN: 0975-3087 & E-ISSN: 0975-9115, Volume 6, Issue 1, 2015, pp.-294-304.
Available online at http://www.bioinfopublication.org/jouarchive.php?opt=&jouid=BIJ0000001

THREE DIMENSIONAL STRUCTURE PREDICTION AND *IN SILICO* FUNCTIONAL ANALYSIS OF GAMMA TOCOPHEROL METHYL TRANSFERASE FROM *Glycine max*

VINUTHA T.[1], BANSAL N.[1], PRASHAT G.R.[2], KRISHNAN V.[1], KUMARI S.[1], DAHUJA A.[1], SACHDEV A.[1] AND RAI R.D.[1]*

[1]Division of Biochemistry, Indian Agricultural Research Institute, New Delhi- 110 012, India.
[2]Division of Genetics, Indian Agricultural Research Institute, New Delhi- 110 012, India.
*Corresponding Author: Email- rajdrai@gmail.com

Received: November 03, 2014; Revised: December 09, 2014; Accepted: January 01, 2015

Abstract- γ-Tocopherol methyl transferase (γ-TMT) involved in synthesis of tocopherol (vitamin-E) methylates γ- and δ- tocopherols to form α - and β-tocopherols respectively. γ-TMT of soybean (*Glycine max* L.) was found to be a water soluble protein with the highest amount of polar amino acids contributing to its molecular surface hydrophilicity; the protein was found to be highly thermostable with half-life of < 5hr under *in-vivo* conditions. The isoelecric point (6.3) of the protein rendered the protein activity in the acidic buffer. We predicted three dimensional structure of γ-TMT as a monomer harboring majority of the α-helical structures and with the highest amount of hydrogen-bonded turns and extended strands in the β-ladder. Qualitative and quantitative analyses of the resulting model suggested the proposed model to be reliable with MPQS value of 1.24, an estimated native overlap at 3.5A^0 of about 72.5%, a discrete optimized protein energy of - 0.48 and with a Z-score of 51.10. The predicted model was found to be stable taking into consideration more than 94.1% of the residues in the most favored regions. The structural superimposition of the predicted structure indicated a highly conserved structure despite its low amino acid similarity with the template protein. The results also led to the identification of the functional SAM/SAH binding sites such as HIS38, HIS40, GLY88 and ILE111 on γ-TMT and revealed the presence of the largest cleft on the surface which may play a major role during the ligand-protein interactions. Phylogenetic tree analysis revealed the *Glycine max* γ-TMT to be evolutionarily modified from photosynthetic bacterial MPBQ methyl transferase. Thus the predicted three dimensional structure and other related information generated in the present study have potential implications in better understanding of the molecular mechanisms and nature of methyl transferase related enzymatic reactions.

Keywords- *Glycine max*, γ-TMT, secondary structure, ligand binding site, SAM, Motif

Citation: Vinutha T., et al. (2015) Three Dimensional Structure Prediction and *In Silico* Functional Analysis of Gamma Tocopherol Methyl Transferase from *Glycine max*. International Journal of Bioinformatics Research, ISSN: 0975-3087 & E-ISSN: 0975-9115, Volume 6, Issue 1, pp.-294-304.

Introduction

Methyl transferases (EC 2.1.1) constitute an important class of enzymes present in every life form and transfer a methyl group, most frequently from S-adenosyl-L-methionine (SAM or Adomet), to a nucleophilic acceptor, such as oxygen, leading to S-adenosyl-L-homocysteine (AdoHcy) and a methylated molecule. These enzymes have in common a conserved region of about 130 amino acid residues that allow them to bind SAM [1]. The substrates that are methylated by these enzymes cover virtually every kind of biomolecule ranging from the small molecules to lipids, proteins and nucleic acids [2,3]. Methyl transferases are therefore involved in many essential cellular processes including biosynthesis, signal transduction, protein repair, chromatin regulation and gene silencing etc. [1]. More than 230 families of methyl transferases have so far been described, of which more than 220 use SAM as methyl donor [4].

Gamma- tocopherol methyl transferase, (EC 2.1.1.95) also called tocopherol O- methyl transferase, belongs to the class methyltransferases (2.1.1) and is involved in the synthesis of tocopherols (vitamin- E). It methylates γ- and δ-tocopherols to form α- and β-tocopherols respectively. Tocopherols with potent antioxidant properties are synthesized by photosynthetic organisms and play very important role in human and animal nutrition. In soybean, γ-tocopherol is the predominant form found in the seeds, whereas α-tocopherol is the most bioactive form, suggesting that the final step of α-tocopherol biosynthetic pathway catalyzed by γ-TMT is a limiting one [5]. It has been observed that the expression of γ-TMT varies within the soybean varieties with α-tocopherol content variation between 20 to 30% [6]. Thus greater understanding of the molecular structure of γ-TMT protein can provide a deeper insight into molecular processes related to wide variations in the accumulation of α -tocopherol which is an important γ-TMT catalyzed end product. α-Tocopherol also play an important role in the photosynthesis and macronutrient homeostasis through modulation of signal transduction pathway [7,8].

As X-ray crystallographic structure of γ-TMT has not yet been reported, and also no significant information regarding its catalytic site, functional domain and motif and structural conservation are available, computational tools might prove useful for the researchers to understand these processes through the analysis of physicochemical and structural properties of this protein. Many computational tools are now available for making structural predictions of proteins. Amino acid sequence composition however provides most of the information required for functional characterization of the molecule through its physicochemical properties [9]. In this paper we report homology modeling of γ-TMT proteins from *Glycine max* and *in silico* analysis of its structure. As 3-D structure of this protein is not yet available, these information will be of great help to describe its structural features and to understand its molecular functions. Many computational approaches based on the analysis of protein sequences or structures, have also been developed to predict functional sites, including the ligand binding sites [10] and the cleft & grooves on the surface. Besides elucidating the functional characterization of the protein, potential knowledge of the binding sites can guide to design the inhibitors and antagonists and also provide a scaffold for targeted mutations [11]. In addition, study of evolutionary history, variations in protein sequence and their functions through phylogenetic analysis is considered an important tool in a molecular biologist's bioinformatics tool kit. Such analysis is also essential to understand major evolutionary questions such as the origin and history of macromolecules, developmental mechanisms and phenotypes [12]. Also, the phylogenetic analysis of protein sequence data is integral to protein annotation, function prediction, identification and construction of protein families and protein discovery [13].

Material and Methods

Sequence Retrieval

γ-TMT protein sequence (Glyma09g35680.1) from *Glycine max* was retrieved from phytozome database and searched for homologous sequences using NCBI Blastx programme [Table-1]. The γ-TMT protein (BAK57287.1) of *G.max* was selected for the study.

Table 1- γ-Tocopherol methyltransferase retrieved from the NCBI database

Organism	Sequences identity with *G. max*	Accession No.	Binding Specificity
Lotus japonicus	82%	AAY52459.1	SAM
Cicer arietinum	86%	XP004498827.1	SAM
Morus notabilis	80%	EXB29127.1	SAM
Gossypium hirsutum	78%	ABE41798.1	SAM
Theobroma cacao	76%	XP007029706.1	SAM
Prunus mume	78%	XP008241299.1	SAM
Solanum tuberosum	77%	NP001275191.1	SAM
Solanum lycopersicum	77%	NP001233814.1	SAM
Perilla frutescens	79%	AFP68180.1	SAM
Solanum pennellii	76%	AD224710.1	SAM
Carthamus oxyacanthus	79%	AFO70131.1	SAM
Zea mays	74%	AGF92809.1	SAM
Artemisia Sphaerocephala	76%	ACS34775.1	SAM
Saccharum hyb. Cultivar R570	75%	AGT16736.1	SAM
Triticum aestivum	76%	CAI77219.2	SAM
Brassica napus	75%	ACD03287.1	SAM
Brassica oleracea	76%	AAO13806.1	SAM
Arabidopsis thaliana	76%	NP176677.1	SAM
Helianthus annuus	77%	ABB52800.1	SAM

Phylogenetic Analysis

The phylogenetic tree was generated using the neighbor–joining (NJ) method implemented in MEGA 4.0 software. The alignment of γ-TMT sequences was performed with phylogeny with collapse branches having branch support value smaller than 50% and the NJ tree was bootstrapped by 1000 bootstrap trials to confirm the robustness of the branches. The phylogenetic tree was autogenerated using γ-TMT amino acid sequence as query to analyze all the paralogs of Photosynthetic organisms using KEGG database available at www.genome.jp/kegg. The phylogenetic tree was constructed using NJ method with collapse branches having branch support value < 100%.

Motif Prediction and Analysis

MEME (multiple Em for motif elicitation) which searches via the web server hosted by the National Biomedical Computation Resource (http://meme.nbcr.net) was employed to identify the conserved and novel motifs within the γ-TMT proteins [14]. The search parameters were: motif width: 6-50 amino acids with a maximum 3 motifs for discovery; a motif was considered significant when present in most of the members grouped together in the phylogeny or similar to motif identified in the tocopherol methyl transferases of other plant species.

Protein 3D Structure Prediction

Glycine max γ-TMT structural model was obtained from its amino acid sequence by using MODBASE a queryable database of annotated protein structure models [15]. The predicted structure was revalidated using SWISS-MODEL (http://swissmodel.expasy.org) [16] and protein homology/analog Y recognition engine (PHYRE) (http://www.sbg.bio.ic.ac.uk/phyre/) prediction servers [17]. The ModBase (http://salilab.org/modbase), a database of annotated comparative protein structure models, was used to validate the predicted model as the most reliable model based on the MPQS - (Mod Pipe quality score). This model was selected to predict the secondary structure and the biochemical parameters using PDBsum, a database of mainly pictorial summaries of 3D structures of proteins and nucleic acids in the Protein Data Bank [18] and ProtParam, which computes various physico-chemical properties that can be deduced from a protein sequence [19]. Further, 3-state and 8-state secondary structure for the predicted model was obtained by RaptorX, a protein structure and function prediction server (http://raptorx.uchicago.edu/). The 3-D model obtained was stereochemically evaluated on RAMPAGE server [20] which provides a score based on proline and glycine preferential positions according to the Ramachandran plot; Molecular surface analysis and the conservation of predicted 3-D structure by superimposition with related proteins were studied using chimera software (http://www.cgl.ucsf.edu/chimera) [21]. Ligand–binding sites for the predicted 3-D structure were obtained on 3-D ligand site which is an automated method for the prediction of ligand binding sites [22].

Results and Discussion

Biophysical Characterization of γ-TMT Proteins

The Protparam analyses of all γ-TMT proteins of various plant species included Molecular weight (MW), theoretical PI, instability index, aliphatic index and hydropathicity index which revealed that, the proteins have varying numbers and types of amino acids [Table-2]. Molecular weight of γ-TMT proteins among the given plant species are reported to vary in the range of 30-40 kDa. Isoelectric point (pI) varied between 6.0 to 8.0 with γ-TMT from *Glycine max* showing a pI value of 6.33, indicating that the enzyme is likely to show activity in the acidic buffers. The predicted pI values for γ-TMT proteins shall prove useful for purification purposes through ion-exchange chromatography (IEC) [23]. Our results indicate a range of pI between 6.0 and 8.0 reflecting variation in the length of the γ-TMT proteins from different species and variation in composition of amino acid at N-terminal end across the species [24]. Our results further supported the findings of Khaldi & Shields [24] that similar homologous proteins with pI values between 6.0 and 8.0 cause the elution of proteins at pH values considerably higher than their pI; thus these results might be helpful in eluting γ-TMT proteins from IEC by altering the pH higher than the given pI range.

Table 2- Parameters of δ-Tocopherol methyltransferases of plant species calculated using the Protparam program

Organism	Sequence Length	MW	PI	EC	Ii	Ai	GRAvy	-R	+R
Glycine max	302	33311.2	6.330	65555	41.11	86.29	-0.175	36	34
Lotus japonicus	358	39877.9	5.890	57325	51.81	83.13	-0.091	43	37
Cicer arietinum	362	40062.7	6.520	67420	47.00	81.66	-0.206	41	39
Morus notabilis	357	40049.8	8.610	72795	56.34	80.14	-0.287	38	42
Gossypium hirsutum	344	37811.0	7.100	63410	51.14	79.19	-0.264	35	35
Theobroma cacao	346	38259.7	8.700	64900	54.53	78.70	-0.234	36	41
Prunus mume	353	38426.1	8.010	57910	40.51	86.86	-0.108	37	39
Solanum tuberosum	368	40314.0	8.000	64900	56.24	84.10	-0.217	37	39
Solanum lycopersicum	362	39814.5	8.280	64900	54.53	82.27	-0.229	38	41
Perilla frutescens	297	33171.9	5.930	61795	44.06	80.84	-0.277	39	35
Solanum pennelli	361	39699.3	8.000	64900	54.20	82.22	-0.233	38	40
Carthamus oxyacanthus	300	33177.9	5.710	58815	47.79	85.93	-0.136	38	32
Zea mays	352	38360.0	8.560	63410	55.43	79.12	-0.225	35	39
Artemisia sphaerocephalo	273	30058.5	5.920	57325	44.96	88.64	-0.052	32	28
Saccharum hyb.	354	38518.0	8.540	64900	59.62	79.21	-0.254	36	40
Tritium aestivum	365	39462.9	6.720	67295	52.24	81.42	-0.217	39	38
Brassica napus	347	38242.2	6.720	60305	55.89	88.85	-0.131	38	37
Brassica oleracea	347	38143.8	6.720	58815	58.61	89.42	-0.102	38	37
Arabidopsis thaliana	348	38075.5	6.720	60305	53.08	82.47	-0.164	39	38
Helianthus annus	314	34661.6	5.850	60305	38.26 (stable)	81.75	-0.234	40	35

molecular weight (MW) (g/mol); isoelectric point (PI); extinction co-efficient (EC) (M^{-1} cm^{-1}); instability index (Ii); aliphatic index (Ai); grand average hydropathy (GRAvy); number of negative residues (-R); number of positive residues (+ R).

Difference in the pI of similar γ-TMT proteins from different plant species is also reflective of their wide differences in the signal peptide sequences at the N-terminal regions which could be considered as an essential modification required for the enzyme to target different organelles in the cell. It has also been shown that pI can vary greatly depending on both the insertions and deletions between the orthologs [25], suggesting that varying pI we observed for similar proteins might be useful in studying subcellular localization of the enzyme. Similar results were shown by Khaldi & Shields [25] indicating a shift in the pI of similar proteins may have an impact on the function of the organelle they interact with.

Instability Index

The instability index (Ii) is used to measure *in vivo* half- life of a protein [26]. γ-TMT proteins of the plant species selected in this study showed an instability index ranging between 38.26 to 59.62 [Table-2], with γ-TMT protein from *Glycine max* having an Ii of 41.11. Our results thus suggested that majority of γ-TMT proteins including γ-TMT from *Glycine max* have a half-life of <5 hr. A protein whose instability index is <40 is generally considered to be stable whereas a value >40 predicts it to be of unstable in nature [9,26]. The results thus indicated that the γ-TMT proteins, including the γ-TMT from *Glycine max*, are unstable except that from *Helianthus annuus* with its *in vivo* half-life >16 hr. Stability of protein was estimated from amino acid sequence using commonly used method (expasy proteomics tool), generated by Guruprasad, et al [26] and is based on the correlation between protein stability and its dipeptide composition. The stability of a protein can be represented by a protein instability index score by calculating the average of the dipeptide instability weight values derived from statistical analysis of unstable and stable proteins [27]. Higher stability of γ-TMT proteins from *Helianthus annuus* might suggest its role in higher turnover of α- tocopherol metabolite in the cell which is reflected in very high α-tocopherol content in it (>95% of total tocopherol content) [28,29] in comparison to all other plant species including *Glycine max* [Table-2] which has very low α-tocopherol content (≤10% of total tocopherol content) [6].

Grand Average of Hydropathy (GRAVY)

GRAVY values were determined to provide a view of hydrophobicity of the whole protein. GRAVY indices for γ-TMT protein ranged from -0.052 to -0.287 [Table-2]. The GRAVY values usually vary in the range of ±2, positive scores indicate hydrophobicity and negative scores indicate hydrophilicity [30]. The lower GRAVY value in the present study, indicated that the molecular surface of the γ-TMT protein generally tends to be hydrophilic which is further validated through chimera software version 1.9. The results thus suggest that γ-TMT proteins have greater interaction with water which was further confirmed by *in silico* analysis of all the proteins through SOSUI server which verified them to be water soluble in nature.

Aliphatic Index

Aliphatic indices (AI) of protein measures the relative volume occupied by aliphatic side chains of the amino acids: alanine, valine, leucine and isoleucine. The AI values in the [Table-2] were computationally generated based on the AI of proteins from thermophilic bacteria which was shown to be significantly higher than that of ordinary proteins and hence it can serve as a measure of thermostability of proteins [31,32]. AI of γ-TMT proteins was found to be in the range of 78 to 89 [Table-2] indicating these proteins to be stable over a wide temperature range and to be flexible in nature [33]. Higher AI value for all the selected γ-TMT proteins suggest that these proteins possess higher thermostability which may help in studying relationship between thermostablity of the protein, biological membrane perturbation and abiotic and biotic stresses often faced by the plant. Variations among various γ-TMT enzymes with respect to their physicochemical parameters like charged amino acids have been observed [Table-2]. These physicochemical analyses might facilitate development of γ-TMT products which shall not only enhance storability of the product but also expand the temperature range to be applied [34-37].

Secondary Structure Analysis of γ-TMT Proteins

Cys-REC analysis showed that the predicted secondary structures of 16 γ-TMT proteins from different plant species have α-helix, β-sheet and coil structures. The results showed the occurrence of α-helix at higher frequency in all the γ-TMT proteins followed by β–sheet and coils [Table-3]. Detailed analysis of secondary structural elements of γ-TMT from *Glycine max* revealed that 77.5% of amino acids reside are in α-helices, while 65.2% are in β-sheets and 12.3% in coiled coil form [Fig-1](a). Further analysis of γ-TMT structure by PDBSum revealed the presence of super secondary structures [Fig-1](a&b) suggesting its role to serve as good nucleation sites for protein folding [37]. This structural information thus, may help in understanding the relationship between amino acid sequence and the tertiary structure of the of γ-TMT proteins, which in turn, can be used for homology modeling as well as designing of novel proteins based on the structure. The higher percentage of α-helices in γ-TMT proteins revealed that these proteins are more stable based on the hydrogen bonding nature of the α-helices which acts as one of the main forces of secondary structure stabilization in proteins. The topology of γ-TMT proteins from *G. max* [Fig-1](c) showed that, it consists of two structural domains. Domain 1 is constituted of N-terminal and C-terminal ends folded to give α-helical topology, whereas Domain 2 is folded into α/β topology. This finding is supported by earlier reports of Miller, et al [38] and Martin and McMillan [39], where they showed that, methyltransferases, in general, have a bi-domain structure where in the first subdomain contains binding site for methyl group donor, while the second subdomain harbors the binding site for acceptor substrate. Our findings related to topological structure thus, might be helpful in predicting the function of uncharacterized proteins falling under this topological structures [40].

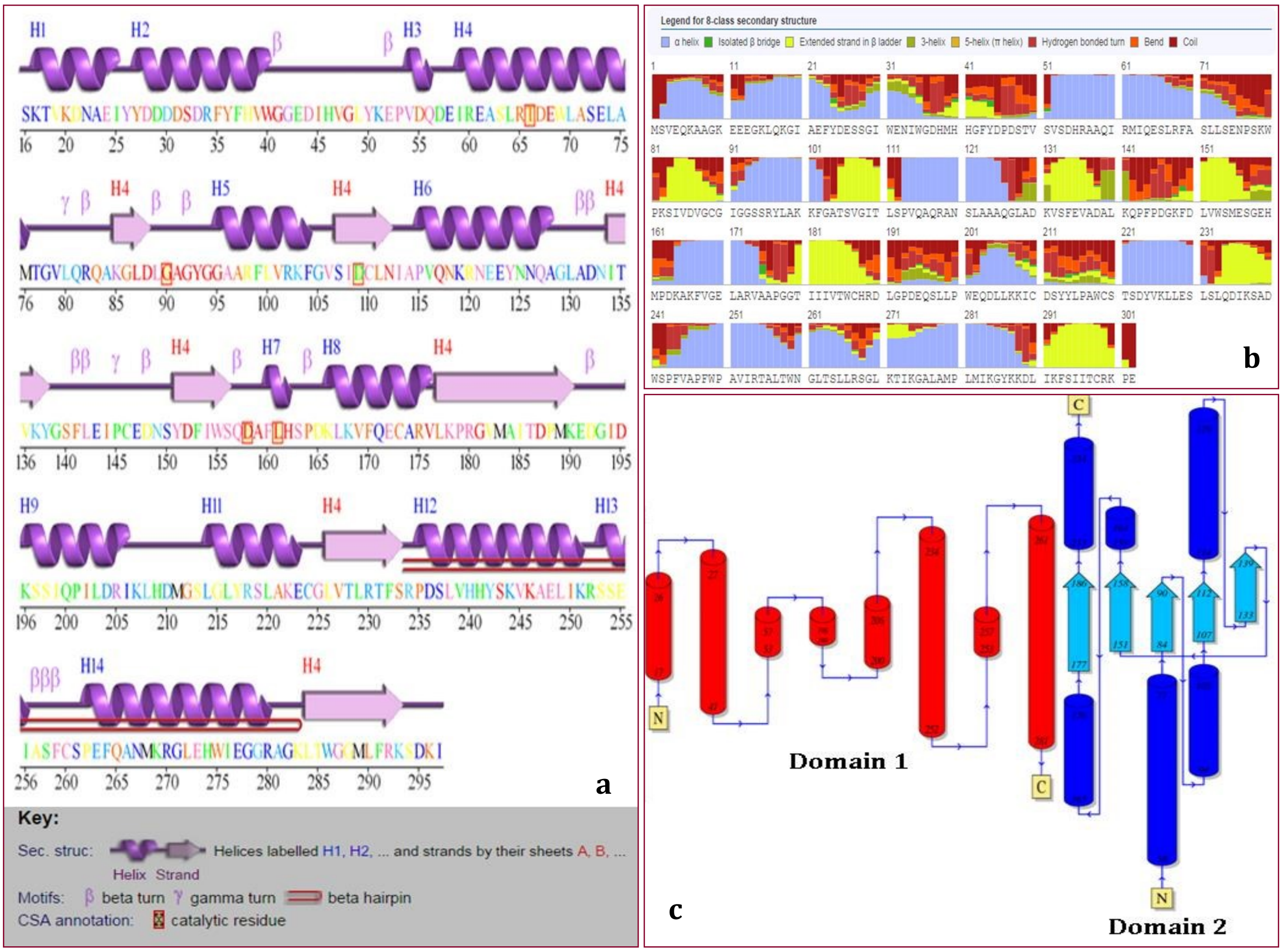
a
H1
H2
H3
H4
H5
H6
H7
H8
H9
H11
H12
H13
H14
Key:
Sec. struc:
Helices labelled H1, H2, ... and strands by their sheets A, B, ...
Helix Strand
Motifs:
β beta turn
γ gamma turn
beta hairpin
CSA annotation:
catalytic residue
b
Legend for 8-class secondary structure
α helix
Isolated β bridge
Extended strand in β ladder
3-helix
5-helix (π helix)
Hydrogen bonded turn
Bend
Coil
MSVEQKAAGK EEEGKLQKGI AEFYDESSGI WENIWGDHMH HGFYDPDSTV SVSDHRAAQI RMIQESLRFA SLLSENPSKW
PKSIVDVGCG IGGSSRYLAK KFGATSVGIT LSPVQAQRAN SLAAAQGLAD KVSFEVADAL KQPFPDGKFD LVWSMESGEH
MPDKAKFVGE LARVAAPGGT IIIVTWCHRD LGPDEQSLLP WEQDLLKKIC DSYYLPAWCS TSDYVKLLES LSLQDIKSAD
WSPFVAPFWP AVIRTALTWN GLTSLLRSGL KTIKGALAMP LMIKGYKKDL IKFSIITCRK PE
c
C
N
Domain 1
Domain 2

Fig. 1- Schematic diagrams showing topology and secondary structure of γ-TMT protein in *Glycine max*: **(a)** α-helices are labeled with the letter "H", and β-strands with the uppercase A, B. β, γ, and hairpin turns are also labeled; **(b)** Eight classes of secondary structures with colour codes are represented; **(c)** Helices are represented as cylinders and β-strands as arrows, the secondary motif map and topology diagram were calculated using the PDBsum tool.

Table 3- Predicted secondary structure and disulfide pattern of γ-TMT proteins. The data was generated from the Protein Sequence Analysis server and CYS REC (http://linux1.softberry.com/berry.phtml).

Organism	α- helix	β-sheet	coil	Disulfide bridge prediction
Glycine max	77.5	65.2	12.3	89-219
Lotus japonicus	87.4	49.2	12	None
Cicer arietinum	79.8	43.4	13.5	6-144, 265-274
Morus notabilis	83.8	41.5	12.3	18-264
Gossypium hirsutum	73	37.5	14	17-261, 148-252
Theobroma cacao	70.8	36.1	14.2	19-263, 37-254
Prunus mume	79.9	43.6	9.9	3-261, 9-270, 27-140
Solanum tuberosum	73.9	35.9	14.1	5-51, 9-18, 253-276
Solanum lycopersicum	76	37.6	13.8	5-44, 18-43, 247-270
Perilla frutescens	76.8	36.7	12.1	None
Solanum pennelli	75.9	37.7	13.9	5-18, 43-246, 44-269
Carthamus oxyacanthus	74.7	36.3	11.7	None
Zea mays	69.2	35.9	14.2	None
Artemisia sphaerocephalo	81	40.3	11.7	None
Saccharum hyb. cultivar R570	66.9	31.4	14.1	None
Tritium aestivum	71	51.2	14.5	None
Brassica napus	82	60.9	11.9	None
Brassica oleracea	78.4	60.2	11.8	255-264
Arabidopsis thaliana	79	58.6	12.1	256-284
Helianthus annus	80.9	43.9	11.8	None

Three Dimensional Structure Prediction of γ-TMT Proteins

The homology model of γ-TMT from *G. max* was generated with the help of MODbase server through Chimera software version 1.9 using putative sarcosine dimethyl glycine methyltransferase from sarcosine *Galdieria sulphuraria* as a template (PDB id 2057) [Fig-2](a). The γ-TMT target region from 16 to 302 amino acid was conserved which showed sequence identity of 22% with the covered template region from 19 to 295 amino acid. Based on the evaluation criteria presented in [Table-4], the γ-TMT protein model was found to be reliable with a probability of >95% for correct folding with the 72.5% of its alpha atoms superpose within 3.5 A^{o} distance from their specific positions [Table-4].

Table 4- Evaluation criteria of a reliable model in comparison with *Glycine max* γ-TMT

Criteria	Evaluation criteria	
	Reliable model of any protein	γ-TMT protein from Glycine max
MPQS (midpipe Quality score)	>=1.1	1.24
TSV mod no 35 (estimated native overlap at 3.5A^{o})	>= 40%	72.50%
GA341^{+}	>= 0.7	0.725
E value	< 0.0001	0
Z Dpoe* (discrete optimized protein energy)	< 0	-0.48

*DOPE is based on an improved reference state that,corresponds to non interacting atoms in a homogenous sphere with the radius dependent on a sample native structure : it thus accounts for the finite and spherical shape of the native structures.

+GA341 is the score for reliability of a model derived from stastical potentials.

Ramachandran plot for γ-TMT of *G. max* was derived by Procheck tool which further validated the reliability of protein's 3-D model; The plot revealed that 94.1% of the amino acid residues were clustered tightly in the most favored regions and only 5.9 % of the residues were scattered in the generally allowed region [Fig-2](a). The result thus showed that, the predicted γ-TMT model by MODbase is validated as a good model. Further the reliability of γ-TMT 3-D structure was confirmed through phyre-2 software wherein, it showed 96% of residues modeled at > 90% confidence.

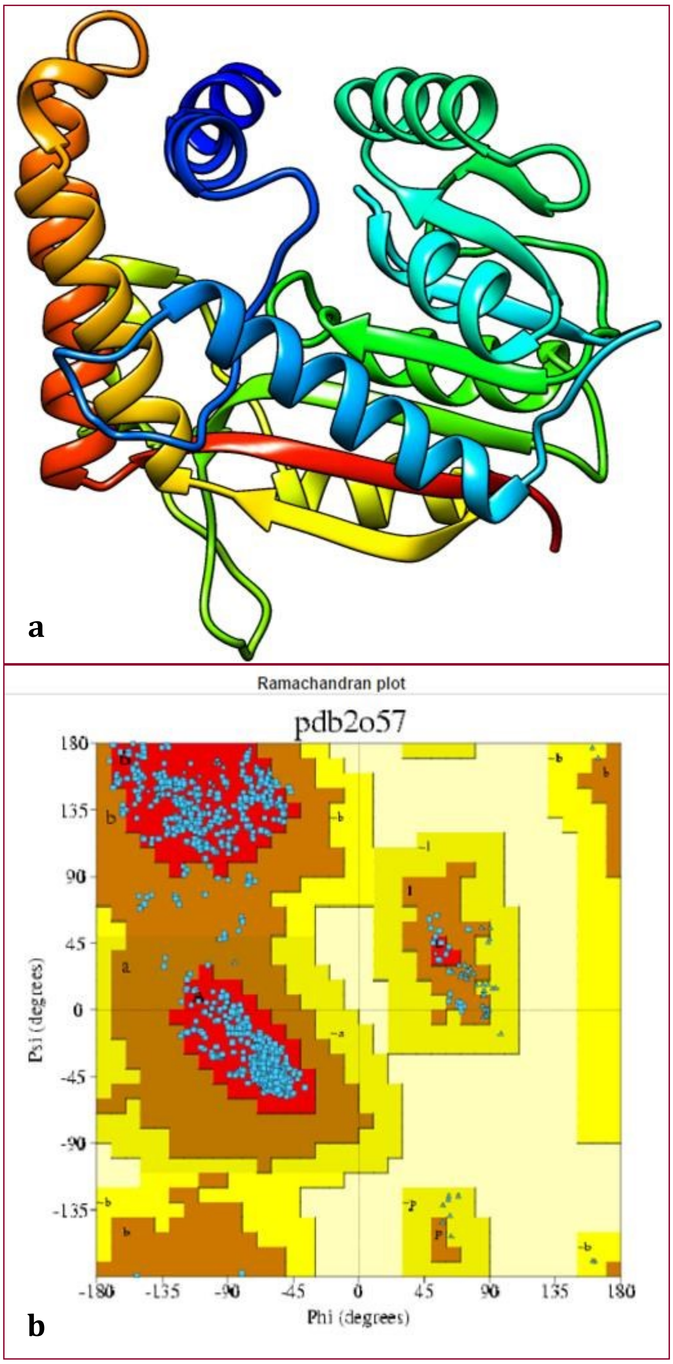

Fig. 2- (a) Predicted 3D structure of the *Glycine max* γ-TMT. The model was generated with Modbase using PDB template 2057. PyMOL was used to visualize the model; **(b)** The Ramachandran plot for the modeled *Glycine max* γ-TMT which was generated with the PROCHECK program; A, B, L letters on the region (coloured red) represents the number of residues in most favoured region.

Characteristics of Molecular Surface of γ-TMT

Amino acid hydrophobicity is a commonly used attribute for the analysis of the molecular surface of a protein molecule. Sructural analysis of γ-TMT using Chimera version 1.9 revealed that most of the amino acid residues are polar in nature (blue), a few of the residues are hydrophobic (orange) and others are in between hydrophobic and polar (white) [Fig-3]. The data thus confirms that γ-TMT protein is highly polar and soluble in nature. The polar nature of the surface suggests that this protein can interact with other subunits of a protein having polar surfaces and hence can serve as a “hot spot” for protein-protein interactions [41].

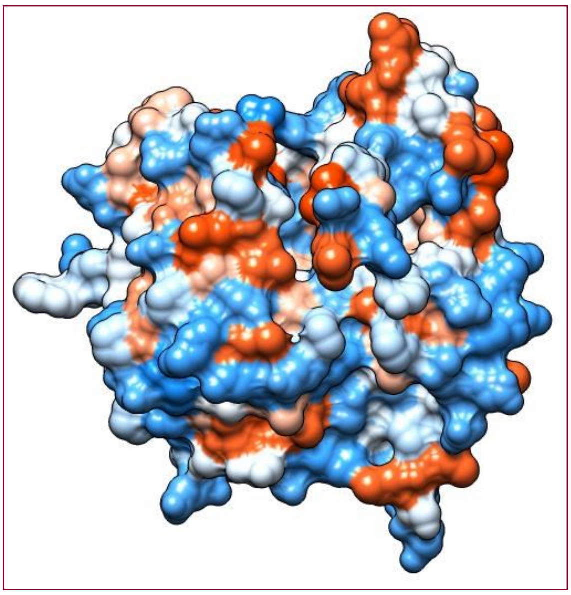

Fig. 3- Molecular surface of *Glycine max* γ-TMT, representing highly polar surface (blue) followed by hydrophobic surface (red).

Structural Homology

The present work demonstrates reliable γ-TMT protein modeling based on the structural homology with the PDB data. Although many functionally diverse methyltransferase structures from different species are available in DALI (Distance Alignment Matrix Method) server, the predicted γ-TMT model showed high structural similarity with PDB code 2057, despite its low (22%) amino acid sequence similarity, as compared to PDB code 2i6gA which showed 38% sequence similarity [Fig-4]. A DALI search for structures similar to γ-TMT returned 903 hits with only first ten different methyltransferases showed Z score of >10 [Table-5] which have differential structural homology with RSM (root mean square) deviation between 0.4 Aº to 3.3 Aº suggesting that the γ-TMT protein has remarkably high structural homology despite its low-sequence identity with the template protein [Table-5], [Fig-4].

Phylogenetic Relationships among γ-TMT Proteins

Phylogenetic tree of 16 γ-TMT proteins showed 2 different groups. Group I included *G. max*, *Arabidopsis thaliana*, *Brassica napus*, *Morus notabilis*, *Lotus japonicus*, *Gossypium hirsutum* and *Theobroma cacao* which were found to descend from *Prunus mume*, whereas Group II included *Helianthus annuus*, *Triticum aestivum*, *Zea mays*, *Saccharum* hybrid cultivar R570, *Artemisia sphaerocephala* and *Carthamus oxycanthus* as descent from *Perilla frutescens* [Fig-5](a). The numbers beside the phylogenetic analysis branches in the [Fig-5](a) represent bootstrap values (> 50%) based on 1000 replications. These results thus depicted that, all the 16 γ-TMT proteins originally descended from two ancestors. Further, to have deeper insight into the γ-TMT protein diversification and its evolutionary relationships, *Glycine max* γ-TMT protein paralogs from all the available photosynthetic organisms on the KEGG database were subjected to phylogenetic tree analysis. The phylogenetic tree revealed that γ-TMT from *Glycine max* originated from evolutionary modification of MPBQ/MSBQ methyltransferase of a photosynthetic bacterium, *Cyanobacterium stanieri* with sequence homology of 35.6% sharing a common SAM binding domain [Fig-5](b). Based on the phylogenetic tree analysis, γ-TMT proteins from plant species were divided into 9 major groups after bootstrapping of branches with neighbor joining distance of 0.1 [6] and compared the amino acid sequence of γ-TMT proteins (γ-TMT 1, 2 and 3) of *G. max* with other plant species, algae and cyanobacteria. This inferred that all γ-TMT isoforms can be classified into one phylogenetic group based on the amino composition. Liscombe, et al [42] studied the relationship between N-methyltransferase candidates from *Catharanthus roseus* with several plant species using phylogenetic tree analysis and showed that, some of the homologous γ-TMT enzymes fall within the clade that includes functionally characterized γ-TMT enzymes. Other N-methyltransferase enzymes fall in a clade representing functionally diverse type I and II methyltransferases including tabersonine 16-O methyltransferase, O methyltransferases of benzylisoquinoline and ipecac alkaloid biosynthesis, and an anthranilate N-methyltransferase. Hu, et al [41] investigated evolutionary relationship among the γ-TMT sequences of selected monocot and dicot plants and found that γ-TMT proteins from the monocots (wheat, rice and maize) were more closely related to each other than to the proteins from other plants under study.

Motif Predictions in γ-TMT

Motif prediction using Motif-based sequence analysis tool - MEME revealed 3 highly conserved motifs; motif I, II and III [Fig-6](a) in γ-TMT protein of *G. max* and other plant species. Interestingly, all the three motifs were found in the same order on the polypeptide chain and were found to be separated at comparable intervals [Fig-6](b) thereby indicating that structural conservation is more important than the amino acid sequence similarity for the protein/enzyme to possess its functional properties. Based on multiple sequence alignment, motif I (starting from 20 to 113) and motif II (staring from 116-245) were identified as methyltransferase-32 domain and motif I and motif II were also been identified as CMAS (mycolic acid cyclopropane synthetase) domain [Fig-6](b). SYSTERS protein family database (http://systers.molgen.mpg.de/) revealed that methyltransferase-32 and CMAS domains belong to the superfamily SAM dependent methyltransferase, suggesting that the highly conserved motif I and motif II have a SAM binding domain which play an important role in methyl group transfer, whereas motif III of γ-TMT protein did not match with any of the annotated domains, except from few plant species viz., G. max, Perilla carthanus, Artemisia sphaerocephalo. Similar kind of results were also shown by Similar results have been reported by Kagan and Clarke [43] indicating that motif I, II and III are commonly found not only in SAM dependent methyl transferases but also in other group of methyltransferases like DNA adenine and cytosine methyltransferases. Six conserved motifs in methyltransferase were reported from *G. max* and *Cicer arietinum* and found that 4 of the motifs have functional catalytic sites in cytosine-5- methyl transferases and 2 were identified as SAM binding sub-domains [44].

Table 5- Structural homologs of γ-TMT using DALI server

pos[a]	Pdb code[b]	prms[c]	rmsd[d]	lali[e]	nres[f]	% id	Description
1	2057-A	46	0.4	274	282	23	Putative sarcosine dimethyl glycine
7	1Lie-B	29	2.3	255	260	15	Mycolic acid synthase
8	3bus-A	28.8	1.9	246	252	26	Methyl transferase
11	4kri-B	28	2.6	248	426	17	Phosphoethanol amine N-methyl transferase2
15	1kpg-c	27.9	2.2	253	285	15	Cyclopropane-fattyacyl-phospholipid synthase-1
30	4f86-5	27.2	2.7	248	273	21	Geranyl diphosphate-2-c-methyl transferase
118	40bw-A	18.2	3.3	186	235	13	2-methoxy-6-polyprenyl-1,4,benzoquinol methylase.
122	35m3-A	17.6	3	117	212	20	SAM- dependent methyl transferases
139	40bx-B	17.2	3.3	188	235	14	2 methoxy-6-polyprenyl-1,4-benzo-quinol methylase
163	2arn-A	16.6	3.3	189	248	19	Ubiquinone/menaquinone biosynthesis methyl transferase

[a] Position in the numerical listing of structural homologs; [b] Z-score; strength of structural similarity in standard deviations above expected; [c] Positional root mean square deviation of superimposed C_ atoms in Å; [d] Total number of equivalenced residues; [e] Length of the entire chain of the equivalent structure; [f] Percentage of sequence identity over equivalent positions.

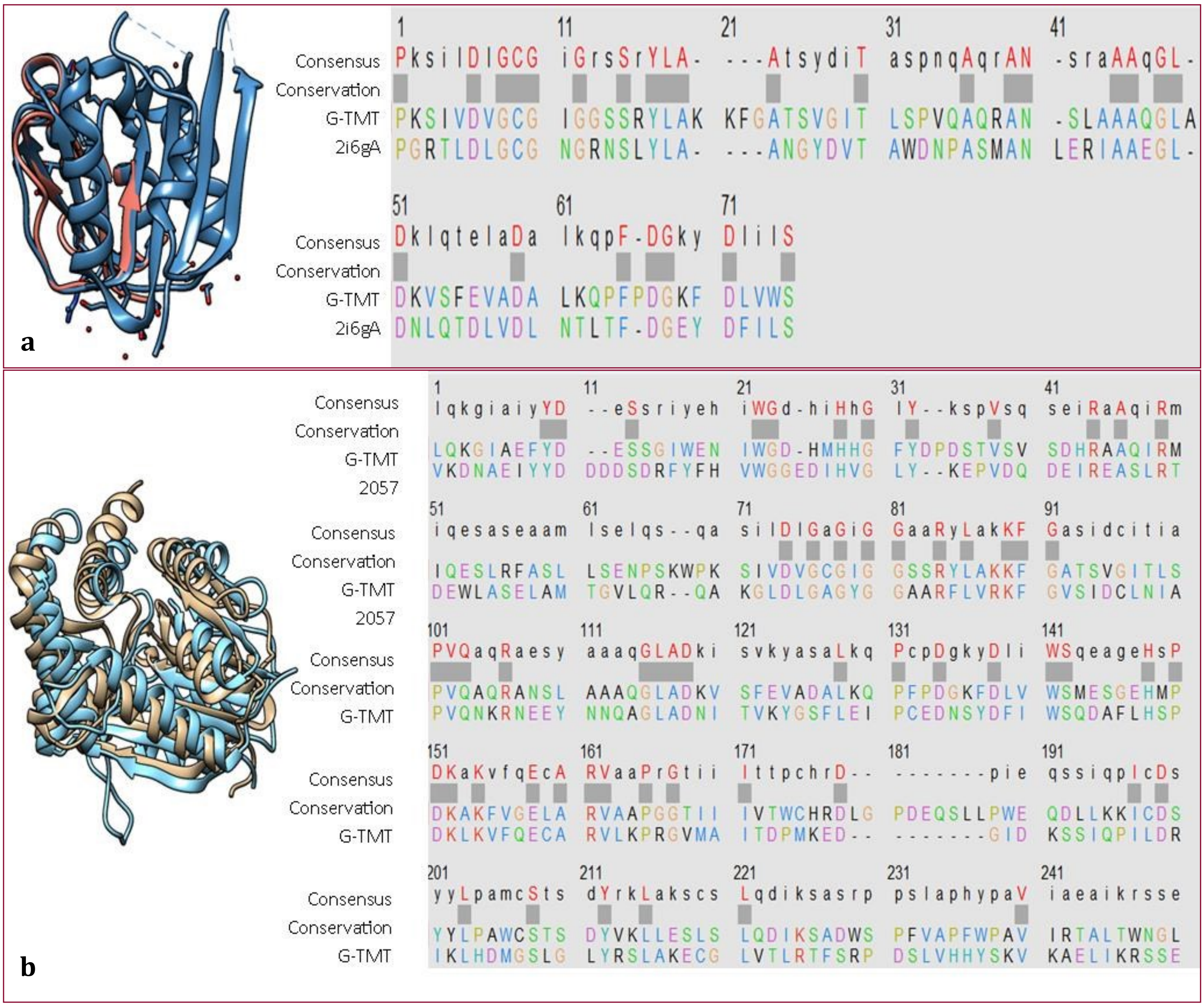
a
Consensus
Conservation
G-TMT
2i6gA
1 11 21 31 41
PksiIDIGCG iGrsSrYLA- ---AtsydiT aspnqAqrAN -sraAAqGL-
PKSIVDVGCG IGGSSRYLAK KFGATSVGIT LSPVQAQRAN -SLAAAQGLA
PGRTLDLGCG NGRNSLYLA- ---ANGYDVT AWDNPASMAN LERIAAEGL-
51 61 71
DklqtelaDa lkqpF-DGky DliIS
DKVSFEVADA LKQPFPDGKF DLVWS
DNLQTDLVDL NTLTF-DGEY DFILS
b
Consensus
Conservation
G-TMT
2057
1 11 21 31 41
lqkgiaiyYD --eSsriyeh iWGd-hiHhG lY--kspVsq seiRaAqiRm
LQKGIAEFYD --ESSGIWEN IWGD-HMHHG FYDPDSTVSV SDHRAAQIRM
VKDNAEIYYD DDDSDRFYFH VWGGEDIHVG LY--KEPVDQ DEIREASLRT
51 61 71 81 91
iqesaseaam lselqs--qa silDIGaGiG GaaRyLakKF Gasidcitia
IQESLRFASL LSENPSKWPK SIVDVGCGIG GSSRYLAKKF GATSVGITLS
DEWLASELAM TGVLQR--QA KGLDLGAGYG GAARFLVRKF GVSIDCLNIA
101 111 121 131 141
PVQaqRaesy aaaqGLADki svkyasaLkq PcpDgkyDli WSqeageHsP
PVQAQRANSL AAAQGLADKV SFEVADALKQ PFPDGKFDLV WSMESGEHMP
PVQNKRNEEY NNQAGLADNI TVKYGSFLEI PCEDNSYDFI WSQDAFLHSP
151 161 171 181 191
DKaKvfqEcA RVaaPrGtii IttpchrD-- -------pie qssiqpIcDs
DKAKFVGELA RVAAPGGTII IVTWCHRDLG PDEQSLLPWE QDLLKKICDS
DKLKVFQECA RVLKPRGVMA ITDPMKED-- -------GID KSSIQPILDR
201 211 221 231 241
yyLpamcSts dYrkLakscs Lqdiksasrp pslaphypaV iaeaikrsse
YYLPAWCSTS DYVKLLESLS LQDIKSADWS PFVAPFWPAV IRTALTWNGL
IKLHDMGSLG LYRSLAKECG LVTLRTFSRP DSLVHHYSKV KAELIKRSSE

Fig. 4- A structural overlay of γ-TMT protein model from *Glycine max*. (a): The structural overlay of γ-TMT protein model from Glycine max (coloured pink) with putative methyltransferase from Salmonella typhimurium (colored dark blue)(PDB code:2i6gA) showing sequence homology (38%) with the γ-TMT from *G.max*; (b): The structural overlay of γ-TMT protein model from Glycine max (coloured yellow) with sarcosine dimethylglycine methyltransferase from Galdieria sulfuraria (PDB id : 2057) (coloured light blue) showing low sequence homology (22%) with γ-TMT from *G.max*. the protein structural graphics were generated from program chimera version 1.9.

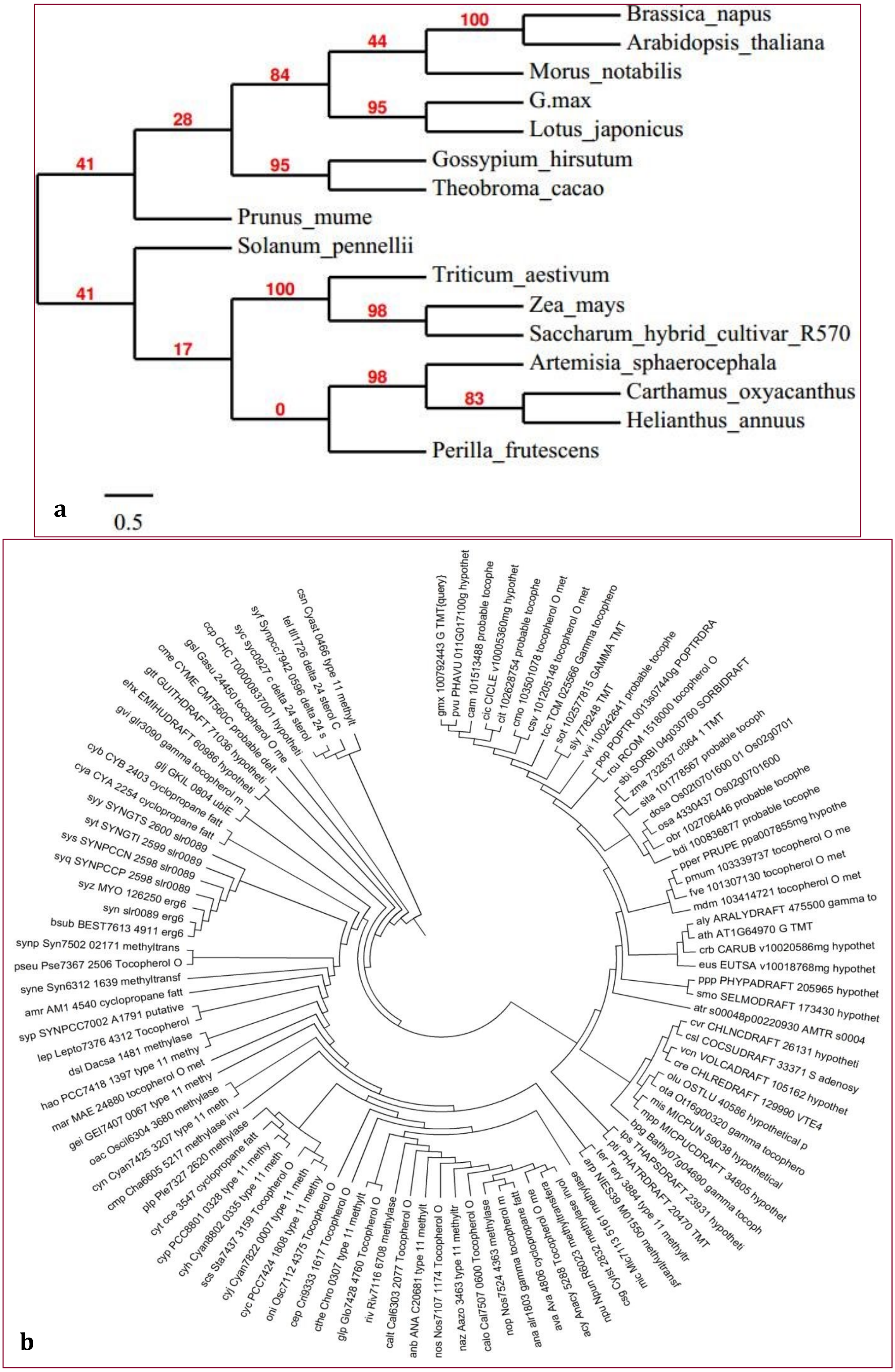
100
44
84
95
28
95
41
41
17
100
98
98
0
83
Brassica_napus
Arabidopsis_thaliana
Morus_notabilis
G.max
Lotus_japonicus
Gossypium_hirsutum
Theobroma_cacao
Prunus_mume
Solanum_pennellii
Triticum_aestivum
Zea_mays
Saccharum_hybrid_cultivar_R570
Artemisia_sphaerocephala
Carthamus_oxyacanthus
Helianthus_annuus
Perilla_frutescens
a
0.5
gmx 100792443 G TMT{query}
pvu PHAVU 011G017100g hypothet
cam 101513488 probable tocophe
cic CICLE v10005360mg hypothet
cit 102628754 probable tocophe
cmo 103501078 tocopherol O met
csv 101205148 tocopherol O met
tcc TCM 025566 Gamma tocophero
sot 102577815 GAMMA TMT
sly 778248 TMT
vvi 100242641 probable tocophe
pop POPTR 0013s07440g POPTRDRA
rcu RCOM 1518000 tocopherol O
sbi SORBI 04g030760 SORBIDRAFT
zma 732837 ci364 1 TMT
sita 101778567 probable tocoph
dosa Os02t0701600 01 Os02g0701
osa 4330437 Os02g0701600
obr 102706446 probable tocophe
bdi 100836877 probable tocophe
pper PRUPE ppa007855mg hypothe
pmum 103339737 tocopherol O me
fve 101307130 tocopherol O met
mdm 103414721 tocopherol O met
aly ARALYDRAFT 475500 gamma to
ath AT1G64970 G TMT
crb CARUB v10020586mg hypothet
eus EUTSA v10018768mg hypothet
ppp PHYPADRAFT 205965 hypothet
smo SELMODRAFT 173430 hypothet
atr s00048p00220930 AMTR s0004
cvr CHLNCDRAFT 26131 hypotheti
csl COCSUDRAFT 33371 S adenosy
vcn VOLCADRAFT 105162 hypothet
cre CHLREDRAFT 129990 VTE4
olu OSTLU 40586 hypothetical p
ota Ot16g00320 gamma tocophero
mis MICPUN 59038 hypothetical
mpp MICPUCDRAFT 34805 hypothet
bpg Bathy07g04690 gamma tocoph
tps THAPSDRAFT 23931 hypotheti
pti PHATRDRAFT 20470 TMT
ter Tery 3884 type 11 methyltr
arp NIES39 M01550 methyltransf
mic MiC7113 5161 methylase invol
csg Cyast 2832 methylase invol
npu Npun R6023 methyltransfera
acy Anacy 5288 Tocopherol O me
ava Ava 4806 cyclopropane fatt
ana alr1803 gamma tocopherol m
nop Nos7524 4363 methylase
calo Cal7507 0600 Tocopherol O
naz Aazo 3463 type 11 methyltr
nos Nos7107 1174 Tocopherol O
anb ANA C20681 type 11 methylt
calt Cal6303 2077 Tocopherol O
riv Riv7116 6708 methylase
glp Glo7428 4760 Tocopherol O
cthe Chro 0307 type 11 methylt
cep Cri9333 1617 Tocopherol O
oni Osc7112 4375 Tocopherol O
cyc PCC7424 1808 type 11 methy
cyj Cyan7822 0007 type 11 meth
scs Sta7437 3159 Tocopherol O
cyh Cyan8802 0335 type 11 meth
cyp PCC8801 0328 type 11 methy
cyt cce 3547 cyclopropane fatt
plp Ple7327 2620 methylase
cmp Cha6605 5217 methylase inv
cyn Cyan7425 3207 type 11 meth
oac Oscil6304 3680 methylase
gei GEI7407 0067 type 11 methy
mar MAE 24880 tocopherol O met
hao PCC7418 1397 type 11 methy
dsl Dacsa 1481 methylase
lep Lepto7376 4312 Tocopherol
syp SYNPCC7002 A1791 putative
amr AM1 4540 cyclopropane fatt
syne Syn6312 1639 methyltransf
pseu Pse7367 2506 Tocopherol O
synp Syn7502 02171 methyltrans
bsub BEST7613 4911 erg6
syn slr0089 erg6
syz MYO 126250 erg6
syq SYNPCCP 2598 slr0089
sys SYNPCCN 2598 slr0089
syt SYNGTI 2599 slr0089
syy SYNGTS 2600 slr0089
cya CYA 2254 cyclopropane fatt
cyb CYB 2403 cyclopropane fatt
glj GKIL 0804 ubiE
gvi glr3090 gamma tocopherol m
ehx EMIHUDRAFT 60986 hypotheti
gtt GUITHDRAFT 71036 hypotheti
cme CYME CMT560C probable delt
gsl Gasu 24450 tocopherol O me
ccp CHC T00008037001 hypotheti
syc syc0927 c delta 24 sterol
syf Synpcc7942 0596 delta 24 s
tel tll1726 delta 24 sterol C
csn Cyast 0466 type 11 methylt
b

Fig. 5- Phylogeny of γ-TMT proteins from selected plant species and other photosynthetic organisms. (a): Phylogenetic tree showing average distance among different γ-TMT proteins. (b): Phylogeny γ-TMT paralogs from photosynthetic organisms available in Kegg database. At the end of each clades short name of the species been written for complete detail see supplementary file.

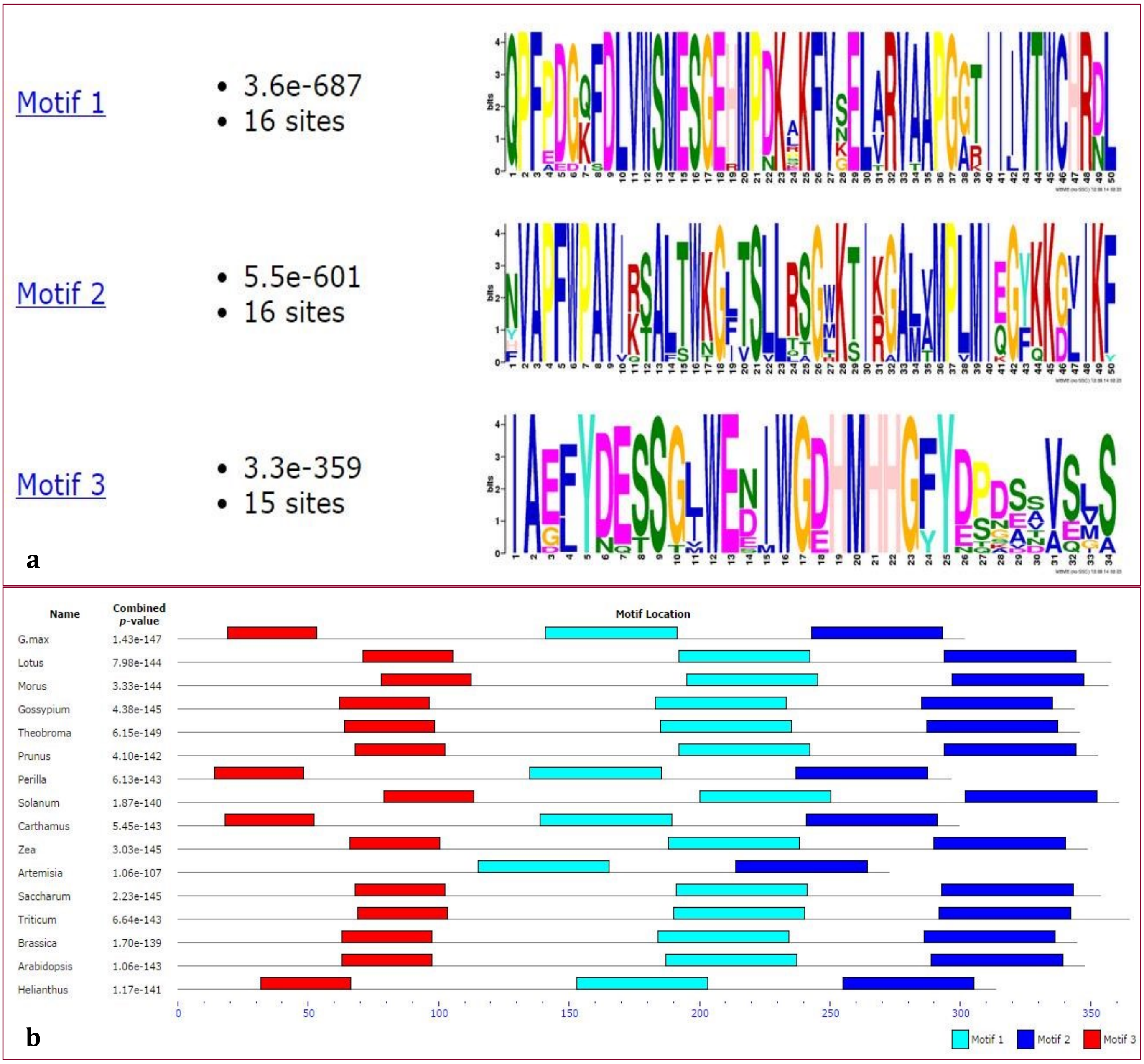

Fig. 6- Motif analysis by MEME; (a). Amino acid sequence logos of motif 1, motif 2 and motif 3; (b). Distribution of conserved motifs in γ-TMT proteins as identified by MEME. The name of each member and combined P value are shown on the left side of the figure. Different motifs are indicated with different colour boxes.

Prediction of Ligand Binding Site and Cleft Analysis of γ-TMT

A ligand binding site prediction server, 3-D ligand site (www.sbg.bio.ic.ac.uk/3dligandsites), was used to determine ligand binding site for γ-TMT protein using its amino acid sequence. It was observed that SAM (S-adenosyl-methionine) and SAH (S-adenosyl-homocysteine) fit into the predicted binding sites of γ-TMT [Fig-7]. The results revealed that approximately 25 SAM or SAH molecules can bind to the residues viz. HIS38, HIS40, GLY88 and LEU111 with an average separation distance of 0A˚ between the substrate and ligand which implies that these ligands could be in physical contact with binding site, suggesting HIS38, HIS40, GLY88 and LEU111 as potent cofactors or inhibitor binding sites. It was also clear that all these residues come under motif I which is reported as SAM binding domain [43], suggesting thereby that SAM binding sites are conserved across the species. These predicted protein binding sites could be of immense importance in assessing 'loss of function' effects on the phenotype of the plant species through mutational studies at the SAM/SAH binding site. Cleft analysis of γ-TMT using Profunc server [45] http://www.ebi.ac.uk/thornton-srv/database/profunc/), predicted clefts and grooves on the protein surface, suggesting a region with average depth of 19.76Aº, accessible vertices 69.44% and buried vertices 14.37% as the large and deepest cleft in this protein [Fig-7](c). Average depth of the cleft i.e. 19.76Aº indicated that the accessibility to the external surface of the domain is restricted. The largest cleft (red portion in [Fig-7](c)) of γ-TMT could act as an active site of the protein as also reported by Sefid, et al [46] for the membrane protein BauA (Baumanniia cinetobactin utilization) from *Acinetobacter baumannii* pathogen.

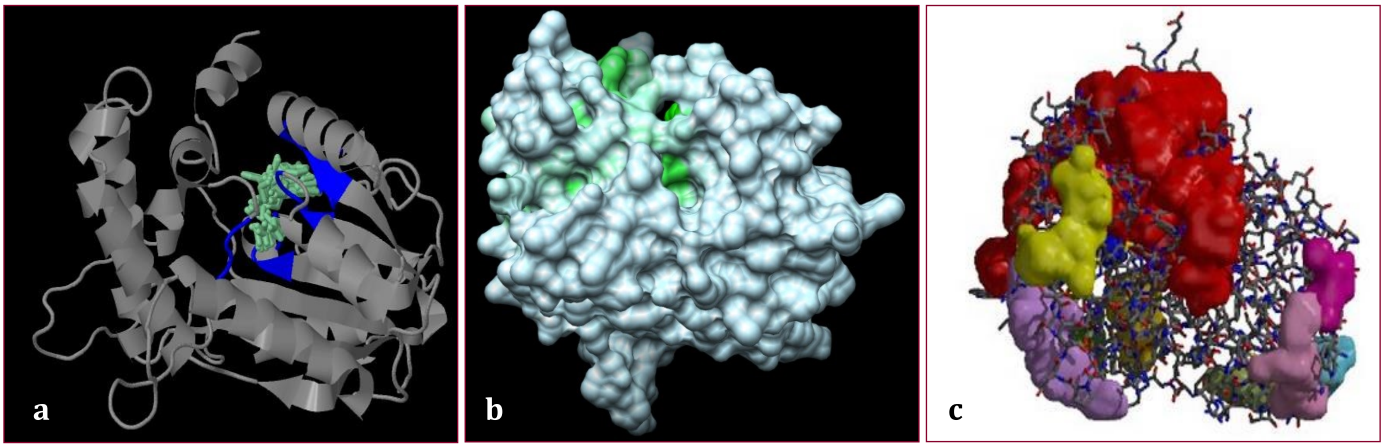

Fig. 7- γ-TMT ligand binding sites predicted by 3Dligandsite server, **(a)**: γ-TMT structure in contact with SAM/SAH ligand ((blue); **(b)**: γ-TMT is shown in ribbon and ligand in the space filling model and the molecular surface of the γ-TMT protein showing largely the poalr or hydrophilic surface (grey); **(c)**: Molecular surface of γ-TMT showing cleft and cavities in protein structure, the largest cleft shown in red consists of negatively charged amino acids (aspartate, glutamate).

Conclusion

We have predicted three dimensional structure and physicochemical properties of γ-TMT of *Glycine max* deploying the most reliable computational tools. The study indicates γ-TMT enzyzme to be highly polar, soluble in nature and possess optimum activity in acidic buffer. Although γ-TMT from *G. max* was found to be highly thermostable, it was predicted to be unstable under *in vivo* conditions. Overlaying of γ-TMT conformations with other similar proteins (PDB id: 2057 and 2i6gA) showed the importance of structural similarity over amino acid similarity for the function of the protein and its activity. Structural analysis revealed high frequency of α-helices and HIS38, HIS40, GLY88 and LEU111 to be in the ligand binding sites, conserved within the active site of an enzyme. This was further supported by the presence of three conserved motifs, among which motif I and II showing methyltransferase activity. Phylogenetic tree analysis revealed that, γ-TMT of *G. max* seems to have originated from photosynthetic bacteria *Cyanobacterium stanieri*. Thus understanding the origin of γ-TMT will provide new insights into the transfer of genetic information among the three domains of life and other information generated in this study may help in better understanding the molecular functions and structural properties of γ-TMT of soybean specially its role in tocopherol biosynthetic pathway.

Acknowledgements: This work is supported by grants (SB/YS/LS-100/2013) from Science and Engineering Research Board (SERB) - a statutory body under Department of Science & Technology, Government of India and Indian Council of Agricultural Research.

Conflicts of Interest: None declared.

References

[1] Schluckebier G., O'Gara M., Saenger W. & Cheng X. (1995) *Journal of Molecular Biology*, 247(1), 16-20.

[2] Kozbial P.Z. & Mushegian A.R. (2005) *BMC Structural Biology*, 5, 19.

[3] Wlodarski T., Kutner J., Towpik J., Knizewski L., Rychlewski L., Kudlicki A., Rowicka M., Dziembowski A. & Ginalski K. (2011) *PLoS One*, 6(8), e23168.

[4] Singh S.K., Choudhury S.R., Roy S. & Sengupta D.N. (2008) *Journal of Biomolecular Structural Dynamics*, 26, 235-245.

[5] Tavva V.S., Kim Y.H., Kagan I.A., Dinkins R.D., Kim K.H., Collins G.B. (2007) *Plant Cell Reports*, 26, 61-70.

[6] Dwiyanti M.S., Yamada T., Sato M., Abe J. & Kitamura K. (2011) *BMC Plant Biology*, 11, 152.

[7] Chan A.W.S., Chong K.Y., Martinovich C., Simerly C., Schatten G. (2001) *Science*, 12, 309-312.

[8] Sakurai F., Kawabata K., Koizumi N., Inoue N., Okabe M., Yamaguchi T., Hayakawa T. & Mizuguchi H. (2006) *Gene Therapy*, 13, 1118-1126.

[9] Sahay A. & Shakya M. (2010) *Journal of Proteomics and Bioinformatics*, 3, 48-154.

[10]Capra J.A. & Singh M. (2007) *Bioinformatics*, 23, 875-1882.

[11]Capra J.A., Laskowski R.A., Thornton J.M., Singh M., Funkhouser T.A. (2009) *PLoS Computational Biology*, 5(12), e1000585.

[12]Craig R.A. & Liao L. (2007) *BMC Bioinformatics*, 8, 6.

[13]Rokas A. (2011) *Current Protocols in Molecular Biology*, 19, 96, 19.11.1-19.11.14.

[14]Bailey T.L., Boden M., Buske F.A., Frith M., Grant C.E., Clementi L., Ren J., Li W.W., Noble W.S. (2009) *Nucleic Acids Research*, 37, W202-W208.

[15]Pieper U., Webb B.M., Barkan T.D., Schneidman-Duhovny D., Schlessinger A., Braberg H., Yang Z., Meng C., Pettersen F.E., Huang C.C., Datta S.R., Sampathkumar P., Madhusudhan MS., Sjolander K., Ferrin., Burley K.S. & Sali A. (2011) *Nucleic Acids Research*, 39, 465-474,

[16]Biasini M., Bienert S., Waterhouse A., Arnold K.,, Studer G., Schmidt T., Kiefer F., Cassarino T.G., Bertoni M., Bordoli L. & Schwede T. (2014) *Nucleic Acids Research*, 42, , W252-W258.

[17]Kelley L.A. & Sternberg M.J. (2009) *Nature Protocols*, 4, 363-371.

[18]Laskowski R.A., Chistyakov V.V. & Thornton J.M. (2005) *Nucleic Acids Research*, 33, D266-D268.

[19]Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. (2005) *The Proteomics Protocols Handbook*, Humana Press, 571-607.

[20]Lovell S.C., Davis I.W., Arendal W.B., de Bakker P., Word J.M., Prisant M.G., Richardson J.S. & Richardson D.C. (2003) *Proteins*, 50, 437-450.

[21]Pettersen E.F., Goddard T.D., Huang C.C., Couch G.S., Greenblatt D.M., Meng E.C. & Ferrin T.E. (2004) *Chemistry*, 25(13), 1605-1612.

[22]Wass M.N., Kelley L.A. & Sternberg M.J. (2010) *Nucleic Acids Research*, 38, W469-W473.

[23]Widmann M., Trodler P. & Pleiss J. (2010) *PLoS One*, 5(5), e10546.

[24]Khaldi N. & Shields D.C. (2011) *Biology Direct*, 6, 40.

[25]Mackiewicz P.J.K., Mackiewicz D., Kowalczuk M., Biecek P., Polak N., Smolarczyk K., Dudek R.M. & Cebrat S. (2007) *BMC Genomics*, 8, 163.

[26]Guruprasad K., Reddy B.V. & Pandit M.W. (1990) *Protein Engineering*, 4, 155-161.

[27]Fang L., Green S.R., Baek J.S., Lee S.H., Ellett F., Deer E., Lieschke G.J., Witztum J.L., Tsimikas S. & Miller Y.I. (2011) *Journal of Clinical Investigation*, 121(12), 4861-4869.

[28]Velasco L., Pe´rez-Vich B. & Ferna´ndez-Martı´nez J.M. (2005) *Plant Breeding*, 124, 459-463.

[29]Hass C., Tang S., Leonard S., Traber M., Miller J. & Knapp S. (2006) *Theoritical Applied Genetics*, 113, 767-782.

[30]Kyte J. & Doolittle R.F. (1982) *Journal of Molecular Biology*, 157 (1), 105-132.

[31]Ikai A. (1980) *Journal of Biochemistry*, 88, 1895-1898.

[32]Idicula-Thomas S. & Balaji P.V. (2005) *Protein Science*, 14, 582 -592.

[33]Gupta S.K., Rai A.K., Kanwar S.S., Chand D., Singh N.K. (2012) *Journal of Experimental Botany*, 63, 757-772.

[34]Tian J., Wu N., Chu X. & Fan Y. (2010) *BMC Bioinformatics*, 11, 370.

[35]Potapov V., Cohen M. & Schreiber G. (2009) *Protein Engineering, Design and Selection*, 22, 553-560.

[36]Bae E., Bannen R.M. & Phillips G.N. (2008) *Proc. Natl. Acad. Sci. USA*, 105, 9594-9597.

[37]Fan J. & Nan L. (2007) *Chinese Physics*, 16, 392.

[38]Miller D.J., Ouellette N., Evdokimova E., Savchenko A., Edwards A. & Anderson W.F. (2003) *Protein Science*, 12, 1432-1442.

[39]Martin J.L. & McMillan F.M. (2002) *Current Opinion in Structural Biology*, 12, 783-793.

[40]Bu D., Zhao Y., Cai L., Xue H., Zhu X., Lu H., Zhang J., Sun S., Ling L., Zhang N., Li G. & Chen R. (2003) *Nucleic Acids Research*, 31(9), 2443-2450.

[41]Hu Z., Ma B., Wolfson H., Wolfson H. & Nussinov R. (2000) *Proteins Structure Function and Bioinformatics*, 39(4), 331-342

[42]Liscombe D.K., Usera A.R. & Connor S.E. (2010) *Proc. Natl. Acad. Sci. USA*, 107, 18793-18798.

[43]Kagan R.M. & Clarke S. (1994) *Archives of Biochemistry and Biophysics*, 310, 417-427.

[44]Garg R., Kumari R., Tiwari S. & Goyal S. (2014) *Legumes*, 9(2), e88947.

[45]Laskowski R.A., Moss D.S. & Thornton J.M. (1993) *Journal of Molecular Biology*, 231, 1049-1067.

[46]Sefid F., Rasooli I. & Jahangiri A. (2013) *BioMed Research International*, 172784.