Agric. Sci. Digest., 37(3) 2017: 185-190
Print ISSN:0253-150X / Online ISSN:0976-0547

AGRICULTURAL RESEARCH COMMUNICATION CENTRE
www.arccjournals.com

In silico analysis, structural modeling and phylogenetic analysis of EPSP synthase of *Phaseolus vulgaris*

Meenu Goyal*, Sugandh Chauhan and Pardeep Kumar[1]

Department of Biotechnology, Central University of Haryana,
Mahendergarh 123 031, Haryana, India.
Received: 04-04-2017 Accepted: 24-07-2017 DOI:10.18805/asd.v37i03.8986

ABSTRACT

The 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase is one of the essential enzymes of shikimate pathway and is verified as a specific target of broad-spectrum herbicide glyphosate. It is important to get insights into three-dimensional (3D) structure of this enzyme for engineering new herbicides as well as herbicide resistant crops. The information about 3D structure of EPSP synthase enzyme is not yet available in protein database for any of the plant species. Therefore, we revealed the homology model of *Phaseolus vulgaris* EPSP synthase protein using the structure of EPSP synthase from *E. coli* as template. The resulting model structure was refined by RAMPAGE server, ERRAT and Verify3D. Ramachandran plot analysis showed that conformations for 95.6% of amino acid residues are within the most favoured region. The phylogenetic tree (constructed using MAFFT) separated EPSP synthases of bacteria, monocot and dicot plants into distinct clusters. Our study generated reliable 3D model structure of EPSP synthase in *P. vulgaris*, which can be used for future studies.

Key words: EPSP synthase, Homology modelling, *Phaseolus vulgaris*, Phylogeny, Shikimate pathway.

INTRODUCTION

Shikimate pathway is responsible for the biosynthesis of aromatic amino acids (phenylalanine, tyrosine and tryptophan) and various secondary metabolites (auxin, salicylate, folic acid, phytoalexins, flavonoids etc.). This pathway is essential for survival in case of bacteria, fungi, algae and higher plants but is absent in mammals. Therefore, this pathway is an attractive target for the development of herbicides and antimicrobial agents (Zhou *et al*., 2006). Glyphosate (N-phosphonomethyl glycine) is a successful broad-spectrum herbicide, which has proven as potent and specific inhibitor of 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase. The enzyme EPSP synthase catalyzes the transfer of enolpyruvyl moiety from phosphoenolpyruvate (PEP) to shikimate-3-phosphate (S3P) forming EPSP and inorganic phosphate. Glyphosate inhibits EPSP synthase in a reversible reaction, which is competitive with respect to PEP and uncompetitive with respect to shikimate-3-phosphate (Schonbrunn *et al*., 2001). It is a non-selective herbicide with no soil activity (Sanbagavalli *et al*., 2000).

EPSP synthase has been studied extensively in plants during last few decades because of its importance as herbicide target. Mutagenesis of EPSP synthase was done in various crops to obtain glyphosate tolerant EPSP synthase e.g., proline-106 to serine in *Eleusine indica* (Baerson *et al*., 2002), proline-106 to leucine in *Nicotiana tabacum* (Zhou *et al*., 2006), glycine-100 to alanine in *Agrobacterium* sp. strain CP4 (Funke *et al*., 2006), proline-101 to serine in *N. tabacum* (Comai *et al*., 2012).

Biological function of a protein is often a reflection of its tertiary structure. Structural information for EPSP synthase of plants is not available in Protein Data Bank (PDB), a public repository of protein structures. Gaining insights into the 3D structure of EPSP synthase can help in generating new herbicides for weed control as well as in designing mutagenesis experiments for creating herbicide tolerant EPSP synthase for crop plants. Therefore, the objective of the present study is to generate 3D structure model of EPSP synthase of *P. vulgaris* using homology modeling approach.

MATERIALS AND METHODS

Template selection: The protein sequence of EPSP synthase of *P. vulgaris* (target sequence) was retrieved from the NCBI database (http://www.ncbi.nlm.nih.gov) with accession number ABG88197.1. The template selection was done using BLASTP for the target sequence against PDB database. The best template for target sequence was recognized based on the e-value, percentage sequence identity and percentage sequence coverage.

Physico-chemical characterization: The physico-chemical properties of EPSP synthase sequence like theoretical

*Corresponding author's e-mail: goyalmeenu333@gmail.com
[1]Division of Plant Quarantine, ICAR-National Bureau of Plant Genetic Resources, New Delhi 110 012, India.

isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient (Gill and von Hippel, 1989), instability index (Guruprasad *et al*., 1990) , aliphatic index (Ikai, 1980) and grand average hydropathicity (GRAVY) (Kyte and Doolottle, 1982) were computed using the ExPASy ProtParam server (http://us.expasy.org/tools/protparam.html).

Secondary structure prediction: Secondary structure of a protein defines the arrangement of residues in alpha helix, Pi helix, beta sheets, extended strands or turns. The secondary structure prediction was performed by using SOPMA (Geourjon and Deleage, 1995) server (https://npsa-prabi.ibcp.fr/NPSA/npsa_sopma.html).

Homology modelling: The modeling of the three dimensional structure of the protein was performed by homology modeling server, SWISS-MODEL (http://swissmodel.expasy.org/) (Arnold *et al*., 2006).

Model validation: The model was evaluated basis on geometrical and stereo-chemical constraints using RAMPAGE server (http://mordred.bioc.cam.ac.uk/-rapper/rampage.php), Verify 3D (http://services.mbi.ucla.edu/Verify_3D) (Eisenberg *et al*., 1997) and ERRAT (http://services.mbi.ucla.edu/ERRAT/) (Colovos and Yeates, 1993). The model with least number of residues in the outlier region was selected for further studies.

Phylogenetic analysis: For evaluating the phylogenetic relationships among EPSP synthases of various plants and bacteria (Table 1), the amino acid sequences of selected EPSP synthases were aligned using MAFFT (http://www.genome.jp/tools/mafft/), a multiple sequence alignment program. Rooted phylogenetic tree of EPSP synthase sequences was constructed by using UPGMA method.

RESULTS AND DISCUSSION

Template selection and sequence alignment: The *P. vulgaris* EPSP synthase protein sequence (target sequence) comprises of 522 amino acid residues. According to the result of BLASTP search for the target sequence against PDB, two reference proteins (PDB ID: 3NVS, 1G6S) were identified that showed a high level of sequence identity i.e. 54% and 53% respectively, with the target sequence. Out of these, *E. coli* EPSP synthase (PDB ID: 1G6S) with an e-value of 8e-145 and a query cover of 83% was selected for homology modeling of the target sequence. The sequence alignment of *P. vulgaris* EPSP synthase and *E. coli* EPSP synthase is shown in Fig. 1.

Physico-chemical analysis: The physico-chemical analysis of target and template sequence was performed using ExPASy ProtParam server and the results are shown in Table 2. Isoelectric point (pI) value is the pH at which protein is stable and has no net charge. The computed pI value of EPSP synthase enzyme was less than 7, which indicates that the protein is acidic. The computed isoelectric point will be useful for developing buffer system for enzyme purification by isoelectric focusing method. The total number of negatively charged residues (Asp + Glu, 60) is higher than the total number of positively charged residues (Arg + Lys, 51). The intracellular proteins contain more negatively charged residues (Cedano *et al*., 1997) and it can be suggested that EPSP synthase is intracellular protein. The high extinction coefficient of enzyme indicates the high concentration of Cys, Trp and Tyr residues. The computed extinction coefficient helps in the quantitative study of protein-protein and protein-ligand interactions in solution. The stability of a protein can be estimated by instability index. A protein whose instability index is smaller than 40 is predicted as stable (Guruprasad *et al*., 1990). The instability index value for EPSP synthase was found to be 35.80 indicating the stable nature of the protein. The aliphatic index is defined as the relative volume of a protein occupied by aliphatic side chains. *P. vulgaris* EPSP synthase has very high aliphatic index (94.31). Higher aliphatic index of the enzyme indicates its stability over a wide temperature range.

Structural analysis and model building: The secondary structural features of the target and template sequence were calculated using SOPMA server and the result is shown in Table 3. The EPSP synthase protein is composed of 35.63% random coils, 31.99% alpha helices, 21.07% extended

Table 1: EPSP synthases (with accession number) used in the phylogenetic analysis

EPSP synthase source	Accession No.	Characteristic
E. coli	3FJX	Bacteria
E. coli	1G6S	Bacteria
Vibrio cholerae	3NVS	Bacteria
Arabidopsis thaliana	CAA29828.1	Dicot
Brassica napus	CAA35839.1	Dicot
Capsicum annuum	AEK22121.1	Dicot
Nicotiana tabacum	AAA34071.1	Dicot
Phaseolus vulgaris	ABG88197.1	Dicot
Brachypodium distachyon	XP_003557242.1	Monocot
Eleusine indica	CAD01096.1	Monocot
Oryza sativa	BAF18626.1	Monocot
Sorghum halepense	AGG39654.1	Monocot
Zea mays	XP_008659331.1	Monocot



Fig 1: Sequence alignment of *P. vulgaris* EPSP synthase (target) and *E. coli* EPSP synthase (template)

Table 2: The physicochemical properties computed using ExPASy ProtParam server

Property	*P. vulgaris* EPSP synthase (Target)	1G6S (Template)
Molecular weight	56091.59	46095.7
Theoretical Pi	5.40	5.37
Number of amino acids	522	427
Total no. of negativelycharged residues (Asp +Glu)	60	48
Total no. of positively charged residues (Arg + Lys)	51	38
Extinction coefficient	66390	30745
Instability index	35.80	28.78
Aliphatic index	94.31	94.66
Grand average hydropathicity (GRAVY)	0.042	-0.005

Table 3: Calculated secondary structure elements by SOPMA

Parameter	*P. vulgaris* EPSP synthase (Target)	1G6S (Template)
Alpha helix (Hh)	31.99%	38.88%
3_{10} helix (Gg)	0.00%	0.00%
Pi helix (Ii)	0.00%	0.00%
Beta bridge (Bb)	0.00%	0.00%
Extended strand (Ee)	21.07%	20.61%
Beta turn (Tt)	11.30%	11.48%
Bend region (Ss)	0.00%	0.00%
Random coil (Cc)	35.63%	29.04%
Ambiguous states	0.00%	0.00%
Other states	0.00%	0.00%

strands and 11.30% beta turns in *P. vulgaris*. In case of *E. coli*, the EPSP synthase protein is composed of 38.88% alpha helices, 29.04% random coils, 20.61% extended strands and 11.48% beta turn. Random coils have important functions in proteins for flexibility and conformational changes such as enzymatic turnover (Buxbaum, 2007). The modeling of the three dimensional structure of the protein (Fig. 2) was performed by SWISS-MODEL which is a fully automated protein structure homology-modeling server.

Model validation: The ERRAT analysis for the modelled protein revealed that the overall quality factor of the EPSP synthase of *P. vulgaris* was 93.647, which is making it a

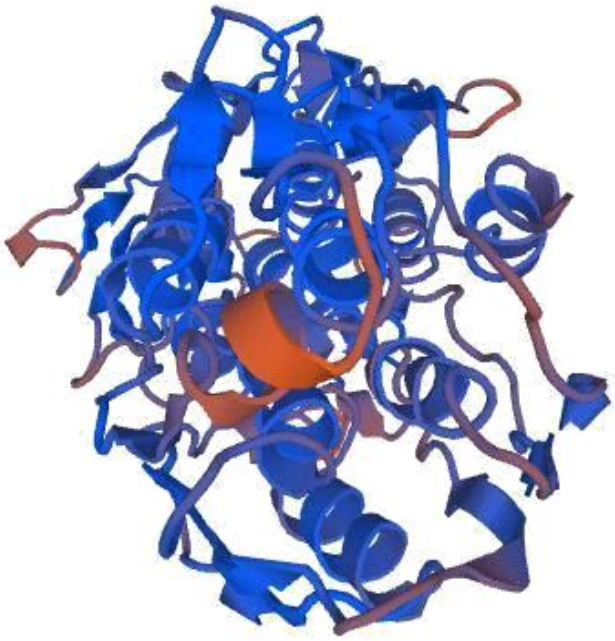

Fig-2: The three-dimensional structure of EPSP synthase of *P. vulgaris* modelled by SWISS-MODEL using *E. coli* EPSP synthase (PDB ID: 1G6S) as template

good quality model (Fig. 3). ERRAT analyzes statistics of non-bonded interactions between different atom types based on characteristic atomic interaction (Colovos and Yeates, 1993). The stereochemical quality of the modelled protein was analyzed by RAMPAGE (Fig. 4). Ramachandran plot analysis showed only 0.7% residues in outlier region, 3.7% in allowed region and 95.6% in favoured region; indicating reliable and good quality model. Verify3D showed a score greater than 0.2 in 100% of the residues that corresponded to a highly acceptable and reliable model. Thus, EPSP synthase model of *P. vulgaris* was generated by SWISS-MODEL server and was confirmed to be reliable and accurate.

Phylogenetic analysis: The phylogenetic analysis of EPSP synthase across the selected organisms showed a clear demarcation of EPSP synthases into two major clusters (Fig. 5). Many of these exhibited orthologous and paralogous relations with each other. Cluster I comprised of sequences from bacteria while cluster II included plant EPSP synthases. Thus during evolution plant EPSP synthases have diversified from the corresponding bacterial enzymes.

Cluster II was further outgrouped into three sub-clusters. Sub-cluster I (*Phaseolus vulgaris, Arabidopsis thaliana* and *Brassica napus*) and II (*Nicotiana tabacum and Capsicum annuum*) included EPSP synthases from dicot plants whereas sub-cluster III comprised of sequences from only monocot (*Brachypodium distachyon, Eleusine indica, Oryza sativa*, *Sorghum halepense* and *Zea mays*) plants. Thus, the distinction of EPSP synthase of *P. vulgaris* (a dicot) from that of monocots and bacteria shows divergent evolution of EPSP synthase in monocots and dicots. Structural and functional importance of this divergent EPSP synthase evolution remains unclear (Garg *et al*., 2014). However, grouping of various EPSP synthases into clusters indicate the presence of conserved domains in EPSP synthase protein gene family during evolution.

CONCLUSION

In the present study, the modeling of the three dimensional structure of the EPSP synthase from *P. vulgaris* was performed by the homology modeling server, SWISS-MODEL. Structural model of EPSP synthase enzyme show reliable folding that was validated by ERRAT and Verify 3D

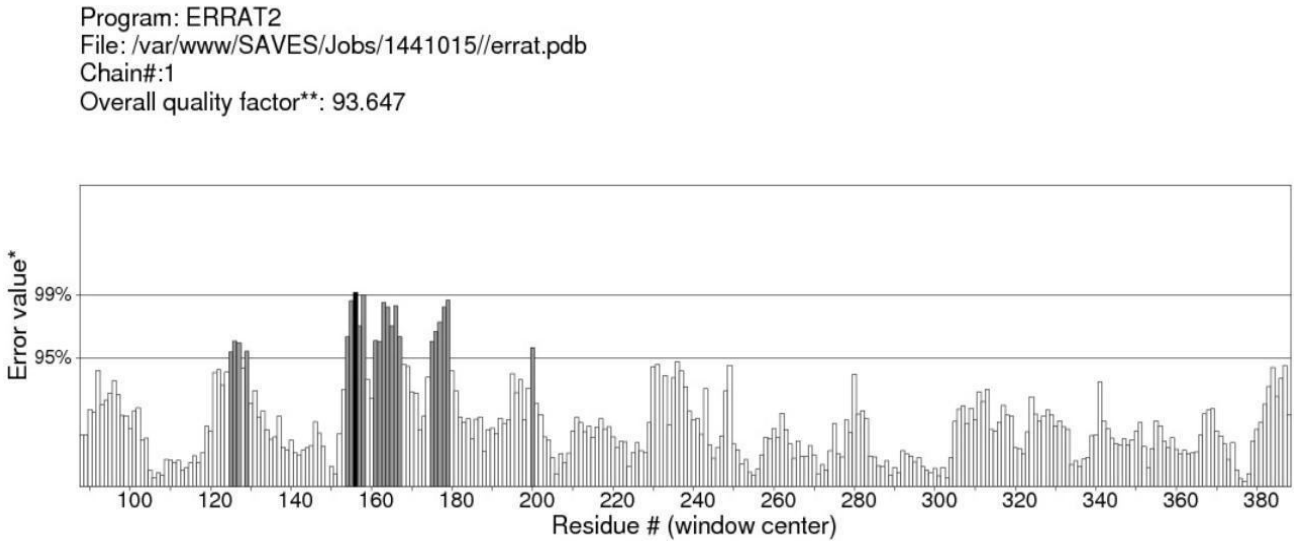

Fig-3: Overall quality factor for the model evaluated by ERRAT

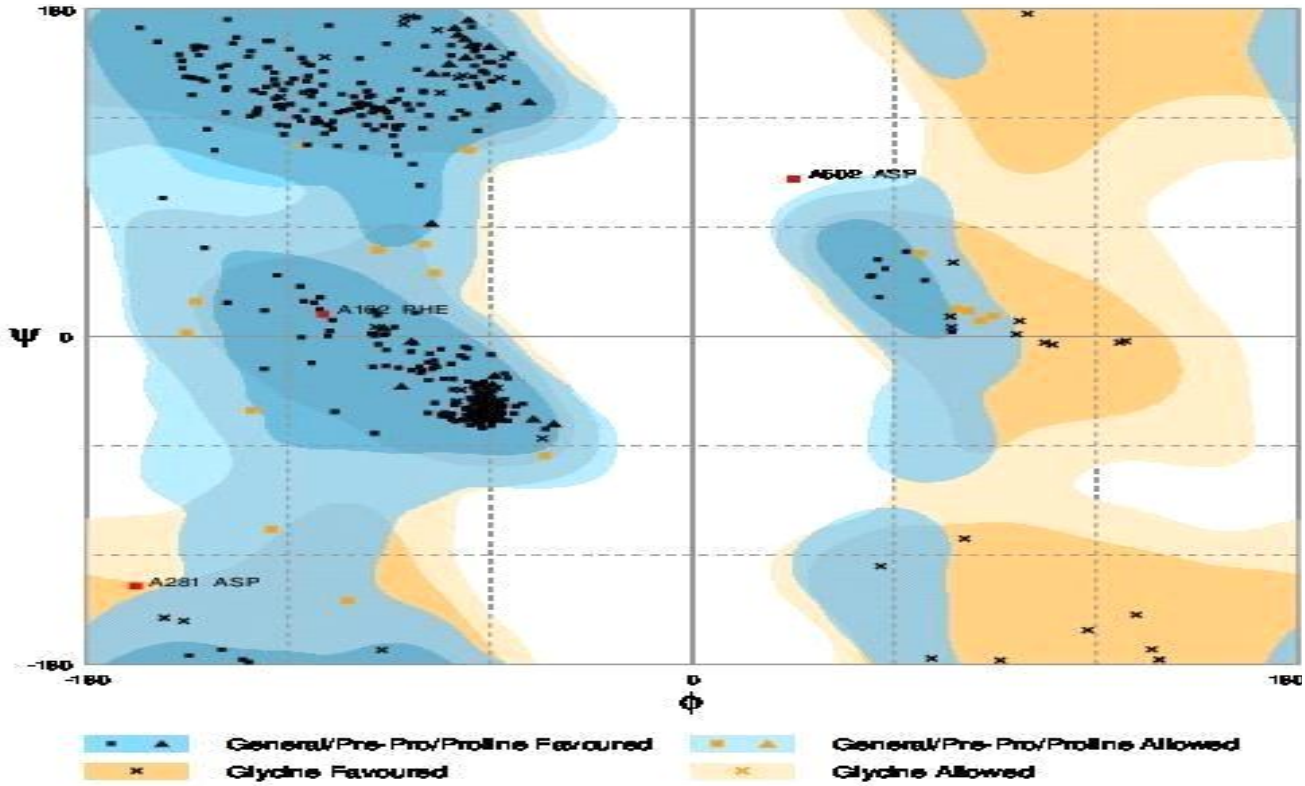

Fig. 4: The plot for *P. vulgaris* EPSP synthase designed by RAMPAGE

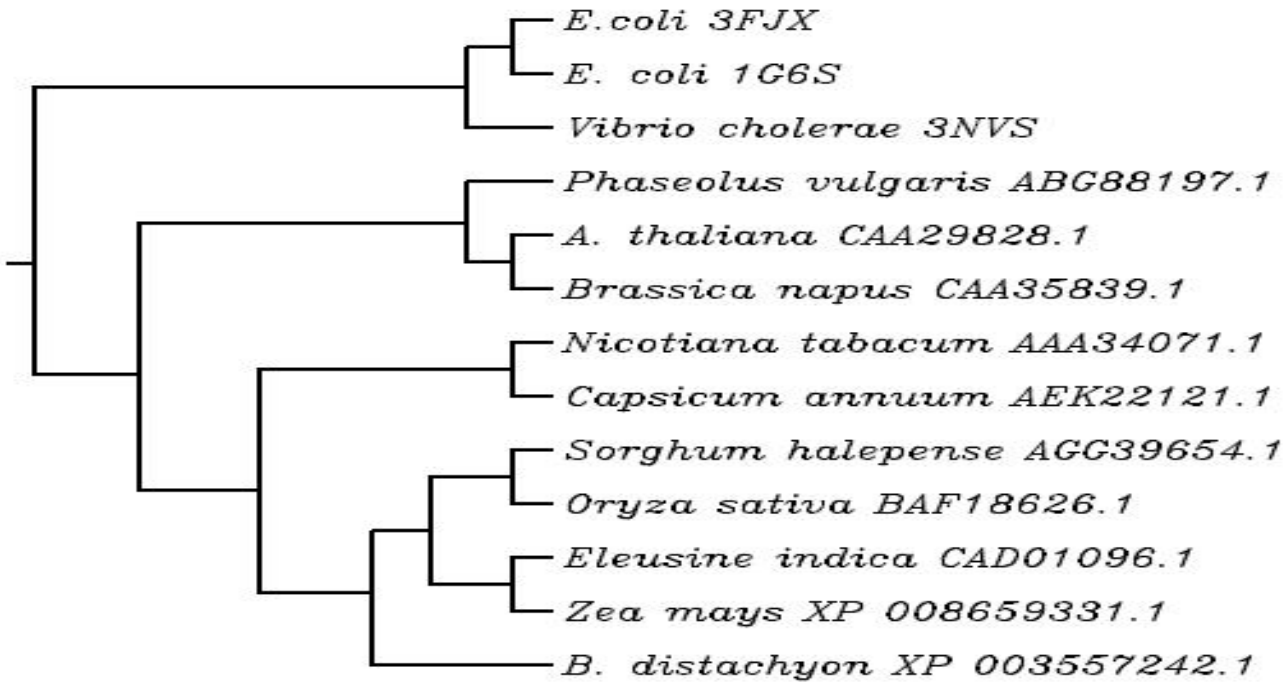

Fig-5: Phylogenetic tree showing relationships among EPSP synthases of various plants and bacteria

servers. This structure can give insights into the active sites of the enzymes and provide a good foundation for functional analysis of the enzyme. Since, EPSP synthase is a direct target of herbicide glyphosate, understanding its fine structure would be a great aid in the development of new herbicides for weed control and herbicide tolerant EPSP synthase for crops.

REFERENCES

Arnold, K., Bordoli, L., Kopp, J., Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling. *Bioinformatics,* **22**: 195-201.

Baerson, S.R., Rodriguez, D.J., Tran, M., Feng, Y., Biest, N.A. *et al.* (2002). Glyphosate resistant goose grass. Identification of a mutation in the target enzyme EPSP synthase. *Plant Physiol.*, **129**: 1265-1275.

Buxbaum, E. (2007). Fundamentals of Protein Structure and Function. Springer Science and Business Media, LLC, New York.

Cedano, J., Aloy, P., Perez-Pons, J.A., Querol, E. (1997). Relation between amino acid composition and cellular location of proteins. *J. Mol. Biol.,* **266**: 594-600.

Colovos, V.C. and Yeates, T.O. (1993). Verification of protein structures: Patterns of non-bonded atomic interactions. *Protein Sci.*, **2**: 1511–1519.

Comai, L., Facciotti, D., Hiatt, W.R., Thompson, G., Rose, R.E. *et al.* (2012). Expression in plants of a mutant aroA gene from *Salmonella typhimurium* confers tolerance to glyphosate. *Nature*, **317**: 741-744.

Eisenberg, D., Luthy, R., Bowie, J.U. (1997). VERIFY3D: Assessment of protein models with three dimensional profiles. *Methods Enzymol.*, **277**: 396–404.

Funke, T., Han, H., Healy-Fried, M.L., Fischer, M. and Schonbrunn, E. (2006). Molecular basis for the herbicide resistance of Roundup Ready crops. *Proc. Natl. Acad. Sci. USA*, **103**: 13010-13015.

Garg, B., Vaid, N., and Tuteja, N. (2014). In-silico analysis and expression profiling implicate diverse role of EPSPS family genes in regulating developmental and metabolic processes. *BMC Research Notes*, **7(1)**: 58.

Geourjon, C. and Deleage, G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biosci.,* **11**: 681-684.

Gill, S.C. and von Hippel, P.H. (1989). Extinction coefficient. *Anal Biochem.*, **182**: 319-328.

Guruprasad, K., Reddy, B.V.P., Pandit, M.W. (1990). Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Prot. Eng.,* **4**: 155-164.

Ikai, A.J. (1980). Thermo stability and aliphatic index of globular proteins. *J. Biochem*., **88**: 1895-1898.

Kyte, J. and Doolottle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.,* **157**: 105-132.

Sanbagavalli, S., Kandasamy, O.S., Ganesan, K. (2000). Herbicide resistance in weeds: a review. *Agric. Rev.*, **21(2)**: 80-88.

Schonbrunn, E., Eschenburg, S., Shuttleworth, W.A., Schloss, J.V., Amrhein, N., Evans, J.N.S., Kabsch, W. (2001). Interaction of the herbicide glyphosate with its target enzyme EPSP synthase in atomic detail. *Proc. Natl. Acad. Sci. USA*, **98**: 1376-1380.

Zhou, M., Xu, H., Wei, X., Ye, Z., Wei, L., Gong, W., Wang, Y., Zhu, Z. (2006). Identification of a resistant mutant of rice 5-enolpyruvylshikimate 3-phosphate synthase using a directed evolution strategy. *Plant Physiol.,* **140**: 184-195.