# Medicinal & Aromatic Plants

**Research Article**         **Open Access**

# *In-silico* Characterization, Structural Modelling, Docking Studies and Phylogenetic Analysis of 5-Enolpyruvylshikimate-3-Phosphate Synthase Gene of *Oryza sativa* L.

Ubaid Yaqoob[1]*, Tanushri Kaul[2], Saurabh Pandey[2] and Irshad Ahmad Nawchoo[1]

[1]Plant Reproductive Biology, Genetic Diversity and Phytochemistry Research Laboratory, Department of Botany, University of Kashmir, Srinagar, Jammu and Kashmir, India

[2]Plant Molecular Biology Lab, International Centre for Genetic Engineering and Biotechnology, New Delhi, India

## Abstract

The 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) is one of the vital enzymes of the shikimate pathway which is involved in the biosynthesis of secondary metabolites and several amino acids. The multiple sequence alignment of these EPSPS protein sequences from different plants showed conserved regions at different stretches with maximum homology in amino acid residues. We revealed the homology model of *Oryza sativa* EPSPS (OsEPSPS) protein using the structure of *E. coli* EPSPS as template. The resulting model structure was refined by PROCHECK, RAMPAGE server, ProSA, Verify3D etc. that indicated the model structure is reliable. Ramachandran plot analysis showed that conformations for 94.3% of amino acid residues are within the most favoured regions. Through motif analysis, it was revealed that a conserved EPSPS domain is uniformly found in all EPSPS proteins irrespective of variable plant species suggesting its possible role in cellular and metabolic functions. The phylogenetic tree constructed revealed different clusters based on EPSPS in respect of bacteria, monocot and dicot plants. The interacting partners of the gene shows the importance of this gene family in regulating developmental and metabolic functions. The two conserved motifs LP(G/S)KSLSNRILLLAAL and LFLGNAGTAMRPL present in almost all EPSPS plant species may function as the catalytic domains of EPSPS enzymes and are supposed to contribute in the glyphosate binding site.

**Keywords:** EPSP synthase; Glyphosate; Herbicide; *Oryza sativa*; Shikimate pathway

## Introduction

The 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS), one of the key enzymes of the shikimate pathway is involved in the biosynthesis of several aromatic amino acids (Phenylalanine (Phe), Tyrosine (Tyr) and Tryptophan (Trp)) and other secondary products (auxin, salicylate, folic acid, phytoalexins, flavonoids, alkaloids etc.) essential for plant survival [1]. It is also verified as a specific target of broad spectrum herbicide glyphosate (N-phosphonomethyl glycine) [2]. EPSPS (aroA) plays a central role in catalysing the transfer of enolpyruvyl moiety from phosphoenol pyruvate (PEP) to shikimate-3-phosphate (S3P) forming EPSP and inorganic phosphate [3]. The reaction is chemically infrequent because it proceeds via C–O bond cleavage of phosphoenol pyruvate rather than via P–O bond cleavage [4]. Glyphosate (GPJ) inhibits EPSPS in a slowly reversible reaction, which is competitive with respect to PEP and uncompetitive with respect to S3P [5,6]. In most of the crops and weeds, glyphosate can starve the plants of aromatic amino acids by competitively inhibiting the binding of EPSPS with PEP. Mutagenesis of EPSPS was done in various species so as to obtain glyphosate-tolerant EPSPS like proline-106 to serine in *E. indica* [7], proline-106 to leucine in *N. tabacum* [1], glycine-100 to alanine in agrobacterium sp. strain CP4 [8], proline-101 to serine in *N. tabacum* [9]. The occurrence of shikimate pathway in algae, bacteria, fungi and plants makes EPSPS a principal target for rising herbicide-resistant genetically modified crops [10]. Thus understanding its mechanism for regulating metabolic and developmental processes in diverse plant species would be a great revolution for engineering new herbicides, developing glyphosate resistant crops, new antibiotic and anti-parasitic drugs.

## Methodology

### Comparative modelling and structural analysis

The reference sequence of EPSPS from *Oryza sativa* was retrieved by using NCBI database (http://www.ncbi.nlm.nih.gov). By searching the PDB of known protein structures, the comparative modelling was performed with target sequence as the query [11]. The target sequence was searched for similar sequence using the BLAST (Basic Local Alignment Search Tool) [12] against Protein Database (PDB) (http://www.rcsb.org). The best template for query sequence was recognized based on the e-value, % sequence identity and % sequence coverage. The BLAST results yielded X-ray structure of EPSPS from *E. coli* with 53% similarity to our target protein (OsEPSPS). Using ClustalW [13], all the sequences of EPSPS were aligned to find out the similarity present among the sequences. 2D and 3D structure alignment was carried out using ClustalW [14] and MATRAS 1.2 [15], respectively. The sequences of the EPSPS were further analysed for the presence of specific EPSPS domains and motifs through motifscan (myhits.isb-sib.ch/cgi-bin/motif scan) and scan prosite (Prosite.expasy.nlm.nih.gov). Analysis of conserved motifs was done by MEME version 3.5.7

**Citation:** Yaqoob U, Kaul T, Pandey S, Nawchoo IA (2016) *In-silico* Characterization, Structural Modelling, Docking Studies and Phylogenetic Analysis of 5-Enolpyrushikimate-3-Phosphate Synthase Gene of *Oryza sativa* L. Med Aromat Plants (Los Angel) 5: 274. doi: 10.4172/2167-0412.1000274

[16] using minimum and maximum motif width of 20 and 50 residues respectively and maximum number of 7 motifs, keeping rest of the considerations at default. Via Modeller 9.12 by comparative modelling of protein structure prediction, the theoretical structure of OsEPSPS from was generated.

The secondary structural features of the EPSPS sequences of template and target were calculated using SOPMA. The physico-chemical properties of EPSPS sequences like molecular weight, theoretical isoelectric point (pI), number of amino acids, total number of positive and negative residues, aliphatic index [17], grand average hydropathy (GRAVY) [18] extinction coefficient [19] and instability index [20] were evaluated by using Expasy's ProtParam server (http://us.expasy.org/tools/protparam.html) [21]. The sub-cellular localizations were predicted by using CELLO v.2.5 [22]. Using NetNglyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/), the N-glycosylation sites of the EPSPS proteins were predicted. Using String software (http://string-db.org/) the interacting partners of EPSPS and its co-expressed genes were predicted [23].

### Model validation of OsEPSPS

On the basis of geometrical and stereo-chemical constraints, the model was evaluated using RAMPAGE server (http://mordred.bioc.cam.ac.uk/-rapper/rampage.php), PROCHECK [24], Verify 3D [25] and ProSA-Web [26]. The model with the least number of residues in the disallowed region was selected for the further studies. The RMSD value between the template and target was calculated using MOE [27]. The best model structure was then compared with the template protein by superimposition using SuperPose Version 1.0 [28].

### Active site prediction and molecular docking

Active sites of model and template proteins were identified using different binding site prediction servers like Q-site finder (http://bmbpcu36.leeds.ac.uk/qsitefinder/), CASTp (http://sts-fw.bioengr.uic.edu/castp/) and PINUP server (http://sparks.informatics.iupui.edu/PINUP/) [29-31]. The refined protein model (OsEPSPS) was used to study its ligand binding mechanism. Docking analysis was performed by Sybyl 8.0 molecular modelling tool to identify active sites on protein structure where favourable protein-ligand interactions can occur [32]. The ligand molecules (S3P and GPJ) were docked inside the cavity of OsEPSPS protein.

### Phylogenetic analysis

Using Molecular Evolutionary Genetic Analysis (MEGA) software Version 4.1 [33], phylogenetic analysis of the sequences was carried by using UPGMA method. Each node was tested using the bootstrap approach by taking 5,000 replicates.

## Results and Discussion

### Comparative modelling and structural analysis

The *Oryza sativa* EPSPS (OsEPSPS) protein sequence comprises of 515 amino acid residues. Sequences that showed maximum identity with high score and low e-value were aligned. According to the result of BLAST search against PDB [34], three reference proteins (PDB ID: 3NVS, 1G6S, 3FJX) represented a high level of sequence identity - 54%, 53% and 53% respectively. The *E. coli* template (PDB ID: 1G6S) with an e-value of 2e-149 and a query cover of 84% was selected for homology modelling. Structurally conserved regions (SCRs) between model OsEPSPS (target) and homologous proteins (PDB: 1G6S, 3NVS, 3FJX) were determined by multiple sequence alignment (Figure 1).

Multiple sequence alignment of the EPSPS sequences highlighted the sequence conservation of amino acid residues among different species (Supplementary File 1). Structurally conserved regions (SCRs) between model OsEPSPS and template (PDB: 1G6S) were also determined (Figure 2). An extensive search of the motifs and their positions was done by MEME software which identified several conserved motifs in the protein sequences of EPSPS (Figure 3). Multilevel consensus sequences for the MEME defined motifs along with their functions are shown in Table 1. LP(G/S)KSLSNRILLLAAL and LFLGNAGTAMRPL motifs were present in almost all selected species.

The initial model of OsEPSPS was built by homology modelling methods using Modeller 9.12 software [35]. The Modeller 9.12 software constructed five model structures for OsEPSPS and the model with the lowest Discrete Optimized Protein Energy (DOPE) score was visualized by Accelrys Discovery studio version 4.1. This model was used for the identification of active sites and for docking of the substrate with the EPSPS. The rice and *E. coli* harbours both of the EPSPS domains which probably indicate toward similar mode of action as in microbes. In this study, predicted 3D structure of OsEPSPS was generated and the N-terminal and C-terminal domains were identified (Figure 4). In *E. coli*, EPSPS consists of six aligned parallel alpha-helices in each of two similar EPSPS I domains [36]. Similar domain structures were detected by Gong et al. [37], Garg et al. [38] and Filiz and Koc [39]. Bacterial EPSPSs are reported to fold in two globular domains and an inside-out α-β barrel domain with PEPS3P binding in the interdomain cleft region [40]. The secondary structural features of the EPSPS sequences of 1G6S and OsEPSPS were calculated using SOPMA [41] with default parameters (Table 2). The EPSPS protein is composed of 42.52% α-helices, 17.86% extended strands and 10.10% beta turn in rice. In case of *E. coli*, the EPSPS protein is composed of 38.88% α-helices, 20.61% extended strands and 11.48% beta turn. Thus the α-helices and the beta sheets cover comparatively larger portions of the rice and *E. coli* EPSPS enzymes. Similar results have been observed by Gong et al. [37], Garg et al. [38] and Filiz and Koc [39] in several plant species. ScanProsite server identified the two signature sequences LFLGNAGTAMRPLTA (166-180) and RVKETERMVAIRTELTKLG (427-445) in both target and template. Several physico-chemical properties of EPSPS sequences were calculated by using Expasy's ProtParam server [21]. The results are shown in Table 3. In developing buffer system for protein purification (isoelectric focusing method), the computed isoelctric point (pI) will be useful. The very high aliphatic index of the EPSPS enzyme sequences indicate that these enzymes may be stable for a wide temperature range. The high extinction coefficient of enzyme in rice indicates the presence of more Cys, Trp and Tyr. The instability index value for the EPSPS proteins were found to be ranging from 28.78 to 33.83 indicating the stable nature of the proteins. Using NetNglyc 1.0 server, the N-glycosylation sites (188 NATY and 464 NITA) of the OsEPSPS protein were predicted and may play role in posttranslational modifications for enzymatic function. N-glycosylation is an essential process for posttranslational modifications of proteins [42].

Using String software, the EPSPS interacting partners as well as its co-expression genes were predicted in both rice and *E. coli* (Figure 5). Some proteins such as 3-dehydroquinate synthase, 3-dehydroquinate dehydratase, shikimate kinase, chorismate synthase and shikimate-5-dehydrogenase are found to be common interacting partners of EPSPS in both rice and *E. coli*. In the second step of shikimate pathway, 3-dehydroquinate synthase converts the 3-deoxy-arabinoheplutosonate-7-phosphate to 3-dehydroquinate and is essential for basic cellular metabolism machinery. In the fifth step of shikimate pathway, Shikimate kinase, an ATP dependent enzyme
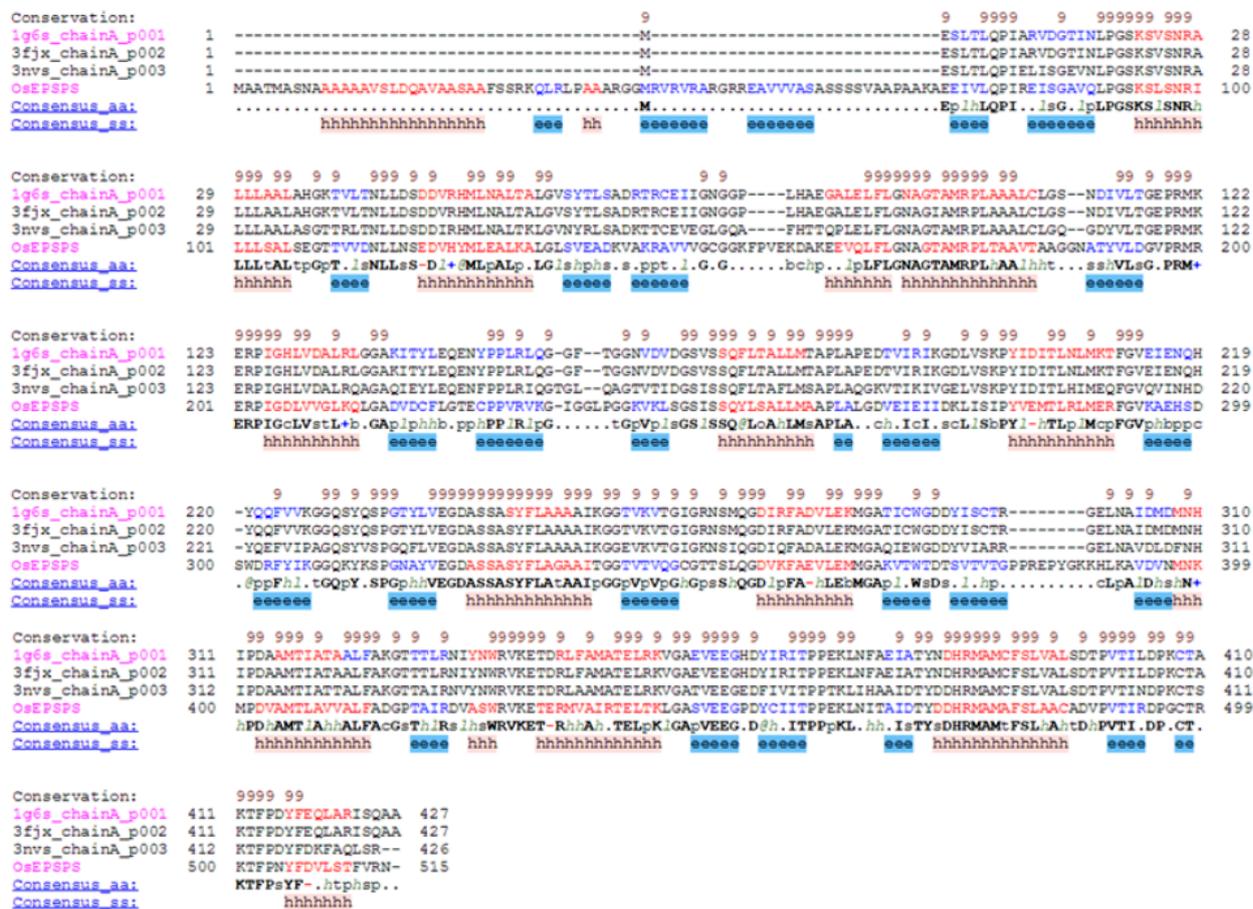
**Figure 1:** Comparative sequences structure alignment of OsEPSPS with other homologues.

**Table 1:** Multilevel consensus sequences for the MEME defined motifs and their predicted functions.

| Motif | Multilevel consensus sequences | Function |
|---|---|---|
| 1 | ITPPEKLNVTEIDTYDDHRMAMCFSLAACADVPVTIKDPGCTRKTFPDYF | Protein kinase C phosphorylation site, Casein kinase II phosphorylation site and N-glycosylation site |
| 2 | DVNMNKMPDVAMTLAVVALFADGPTAIRDVASWRVKETERMIAICTELRK | EPSP synthase, Protein kinase C phosphorylation site. |
| 3 | EGDASSASYFLAGAAITGGTVTVEGCGTNSLQGDVKFAEVLEKMGAKVTW | DLRB LDL-receptor class B (LDLRB), N-myristoylation site |
| 4 | ISSQYLTALLMAAPLALGDVEIEIIDKLISIPYVEMTLKLMERFGVSVEH | Protein Kinase C Phosphorylation Site |
| 5 | VLQPIKEISGTIKLPGSKSLSNRILLLAALSEGTTVVDNLLNSDDIHYML | Casein kinase II phosphorylation site, Protein kinase C phosphorylation site, Pumilio RNA-binding repeat profile. |
| 6 | AVTACGGNARYVLDGVPRMRERPIGDLVDGLKQLGADVDC | EPSP synthase |
| 7 | SWDRFYIKGGQKYKSPGNAYV | - |

**Table 2:** Details of the calculated secondary structure elements by SOPMA.

| Secondary structure element | OsEPSPS | 1G6S |
|---|---|---|
| Alpha helix | 42.52% | 38.88% |
| $3_{10}$ helix | 0.00% | 0.00% |
| Pi helix | 0.00% | 0.00% |
| Beta bridge | 0.00% | 0.00% |
| Extended strand | 17.86% | 20.61% |
| Beta turn | 10.10% | 11.48% |
| Bend region | 0.00% | 0.00% |
| Random coil | 29.51% | 29.04% |
| Ambiguous states | 0.00% | 0.00% |
| Other states | 0.00% | 0.00% |

catalyzes the phosphorylation of shikimate to shikimate 3- phosphate. The seventh step of the shikimate pathway for the biosynthesis of aromatic amino acids is catalysed by chorismate synthase which is conserved in prokaryotes, fungi and plants [43].

### Validation of OsEPSPS structure

RAMPAGE server and PROCHECK generated model revealed that 94.3% residues are falling in the most favoured region, 4.1% residues in allowed region, and 1.6% residues in outlier region of the Ramachandran plot (Figure 6). ProSA-Web analysis of the model revealed a Z-score value of target protein. The Z-score value of the target model OsEPSPS (-8.01) is located within the space of proteins determined by NMR and X-ray crystallography. This Z-score value is close to the value of template 1G6S (-11.83) which suggested that the obtained model was reliable and very close to experimentally determined structures (Figure 7a). Verify3D showed a score greater than 0.2 in 76% of the residues that corresponded to the quality of the OsEPSPS model that was acceptable and reliable. The value of RMSD indicates the degree to which the two three dimensional structures are similar. The lesser the value, the more similar the structures are. The Cα RMSD and backbone RSMD deviation for the OsEPSPS model and the *E. coli* template (1G6S) crystal structure were 1.58Å, and 1.56 Å, respectively and overall RMSD was 1.72 Å. Thus, the OsEPSPS model generated by Modeller 9.12 was confirmed to be reliable and accurate. The superimposition of the template and the model structure is shown in Figure 7b. It shows that the helix and the sheet regions of the template and model structure superimposed in a better way and a large deviation can be observed mainly in loop regions. It is reported that the loop region is the main region where the accuracy of a model protein structure deviates from the templates [44]. The ribbon diagram shown in Figure 4. $^{14}C$ shows the docking of glyphosate (white balls) and S3P (brown balls) into the structure of OsEPSPS (target).

### Prediction of active sites and docking studies

After the final model was built, the possible binding sites of OsEPSPS were searched using various binding site prediction servers such as Q-site finder, CASTp and PINUP [29-31]. These studies showed that residues K, Q, D were highly conserved in active site of both model and the template protein and hence it could be predicted that their biological function would be identical. These conserved residues may function as the catalytic domains of EPSPS enzymes and could be in the glyphosate binding site as seen in bacterial EPSPS [7]. The mutation of a single amino acid (particularly lysine and arginine) can alter the binding site of glyphosate [37]. Molecular docking was performed by Sybyl 8.0 Surflex-Dock method (Tripos Inc., USA). We docked S3P and GPJ inside the cavity of OsEPSPS protein (Figure 7c). The Shikimate-3-phosphate (S3P) has ligand binding residues at 94, 95, 99, 173, 249, 250, 251, 277, 280, 402, 429 and the binding residues are K, S, R, T, S, S, Q, S, Y, D, and K respectively. The glyphosate (GPJ) ligand has ligand binding residues at 94, 170, 172, 202, 251, 402, 430, 433, 474, 475, 500 and the binding residues are K, N, G, R, Q, D, E, R, H, R and K respectively. Both GPJ and S3P have similar amino acids K, Q, D at positions 94, 251 and 402 respectively (Table 4). The glyphosate binding site is dominated by basic residues (Arg and Lys) [45] indicating their role in glyphosate-EPSPS binding.

### Phylogenetic analysis

The phylogenetic analysis of EPSPS across the selected organisms showed a clear delineation of EPSPS into four clusters. Phylogenetic tree results outline the development of EPSPS in *Arabidopsis thaliana, Amborella trichopoda, Brassica rapa, Brachypodium distachyon, Cucumis melo, Fragaria vesca, Glycine max, Malus domestica, Oryza*

---

```
Target      74  EEIVLQPIREISGAVQLPGSKSLSNRILLLSALSEGTTVVDNLLNSEDVH   123
                |.:.||||..:.|.:.||||||:|||.|||:||:.|.||:.|||:|:||.
Template     2  ESLTLQPIARVDGTINLPGSKSVSNRALLLAALAHGKTVLTNLLDSDDVR    51

Target     124  YMLEALKALGLS--VEADKVAKRAVVVGCGGKFPVEKDAKEEVQLFLGNA   171
                :||.||.|||:|   :.||:.  :|:.  |:...::|||||||
Template    52  HMLNALTALGVSYTLSADRT--RCEIIGNGG--PLH--AEGALELFLGNA    95

Target     172  GTAMRPLTAAVTAAGGNATYVLDGVPRMRERPIGDLVVGLKQLGADVDCF   221
                |||||||.||:.   |:...||:.||:|||:||.|.|||.:...||.:..
Template    96  GTAMRPLAAALCL--GSNDIVLTGEPRMKERPIGHLVDALRLGGAKITYL   143

Target     222  LGTECPPVRVKGIGGLPGGKVKLSGSISSQYLSALLMAAPLALGDVEIEI   271
                .....||:|::   ||..||.|.:.:||:|||:|:||||..|..|.|
Template   144  EQENYPPLRLQ--GGFTGGNVDVDGSVSSQFLTALLMTAPLAPEDTVIRI   191

Target     272  IDKLISIPYVEMTLRLMERFGVKAEHSDSWDRFYIKGGQKYKSPGNAYVE   321
                ...|:|.||.:::||.||:.|||:.| :..:.:.|.:||||.|:.|||...||
Template   192  KGDLVSKPYIDITLNLMKTFGVEIE-NQHYQQFVVKGGQSYQSPGTYLVE   240

Target     322  GDASSASYFLAGAAITGGTVTVQGCGTTSLQGDVKFAEVLEMMGAKVTWT   371
                ||||||||||.|||.|||||.|.|.|..|:.|||::||:|||.|||.:.|.
Template   241  GDASSASYFLAAAAIKGGTVKVTGIGRNSMQGDIRFADVLEKMGATICWG   290

Target     372  DTSVTVGPPREPYGKKHLKAVDVNMNKMPDVAMTLAVVALFADGPTAIR   421
                |..::.|        :..|.|:|::|.:||.:||.|||:|.|.|.|.:|
Template   291  DDYISCT--------RGELNAIDMDMNHIPDAAMTIATAALFAKGTTTLR   332

Target     422  DVASWRVKETERMVAIRTELTKLGASVEEGPDYCIITPPEKLNITAIDTY   471
                :::.||||||:..|:.|:.|||.|||:||||:||||||||||:.|.||
Template   333  NIYNWRVKETDRLFAMATELRKVGAEVEEGHDYIRITPPEKLNFAEIATY   382

Target     472  DDHRMAMAFSLAACADVPVTIRDPGCTRKTFPNYFDVLS             510
                :||||||.|||.|.:.|.||||.||.||||.||.:||.:|:
Template   383  NDHRMAMCFSLVALSDTPVTILDPKCTAKTFPDYFEQLA             421
```

**Figure 2:** Comparative sequence alignment of OsEPSPS (target) and *E. coli* EPSPS (template) using superpose.
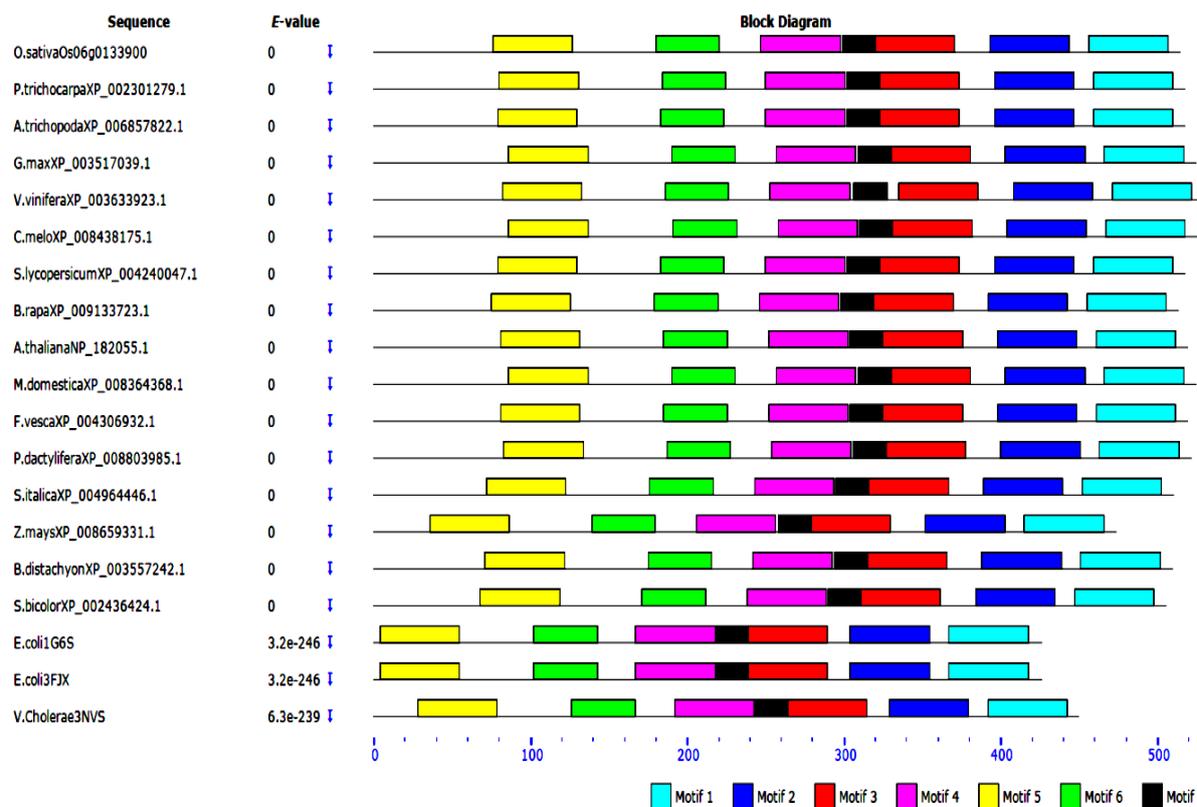
---

**Figure 3:** Block diagram of multilevel consensus sequences for the MEME defined motifs of EPSPS proteins: Seven motifs were obtained by MEME software. Different motifs are indicated by different filled boxes with numbers 1 to 7.
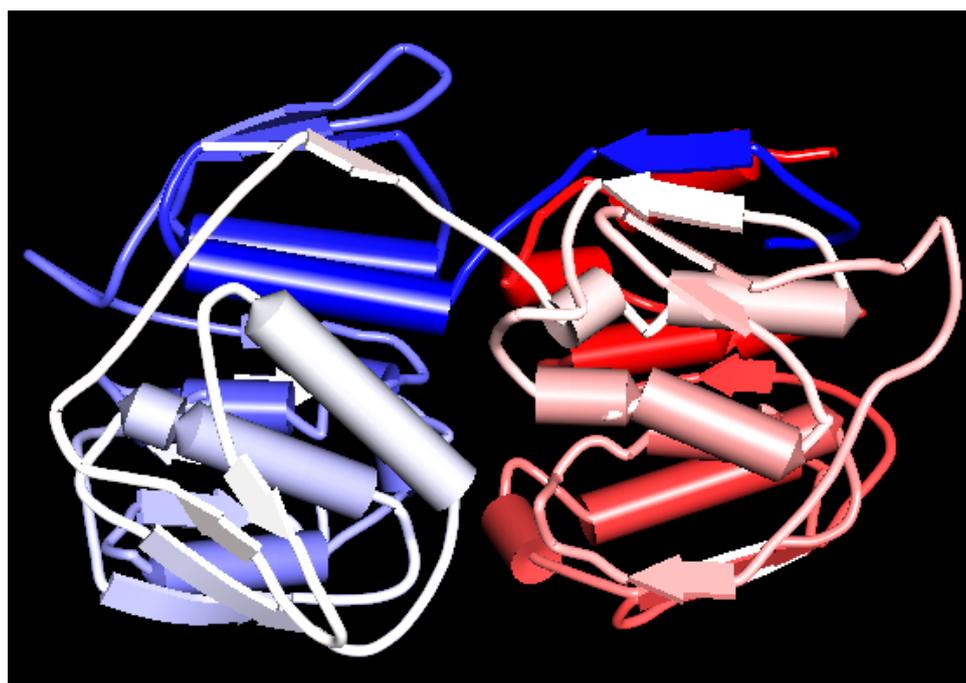


**Figure 4:** Cartoon structure of OsEPSPS showing its N- and C- termini in blue and red respectively.
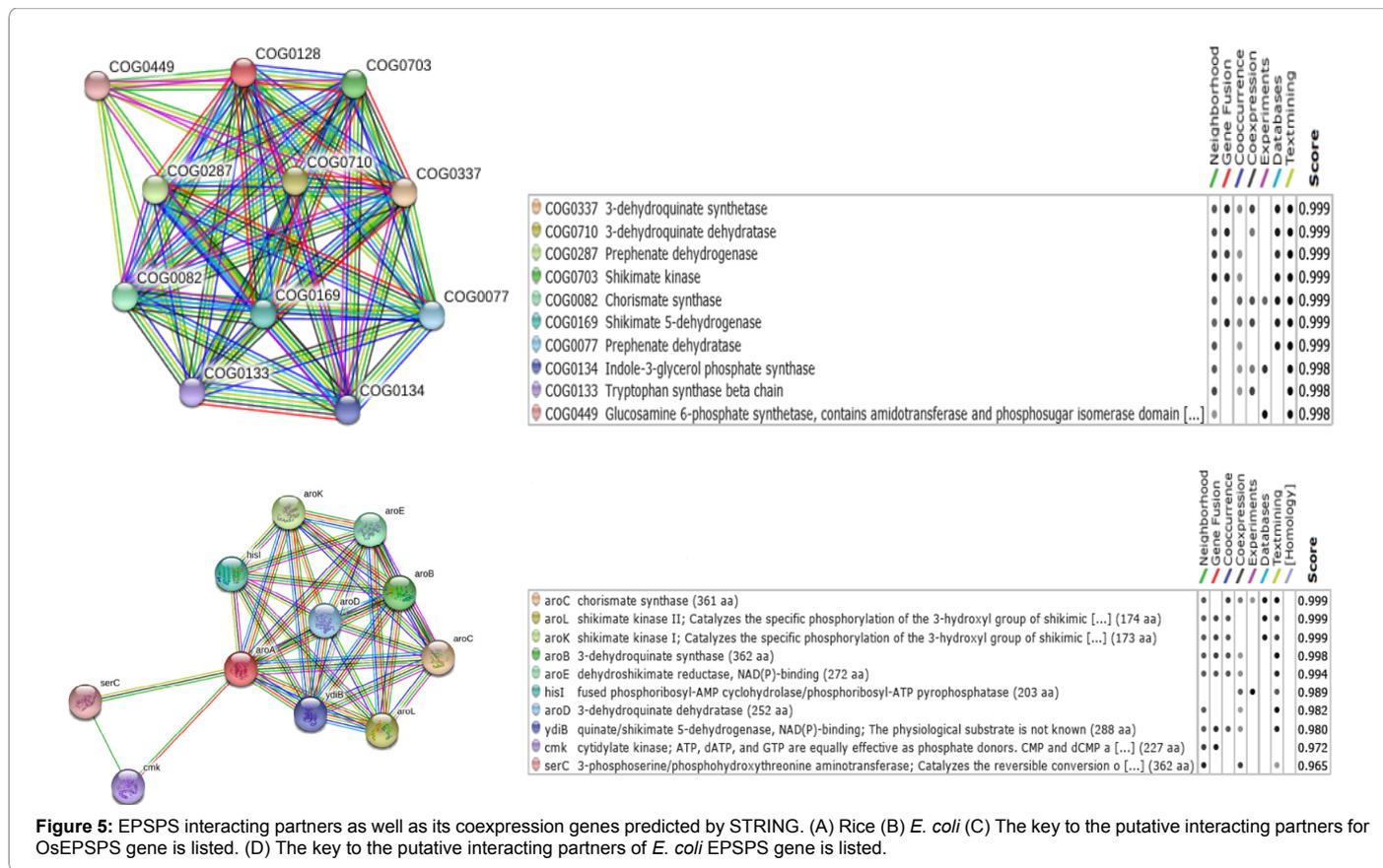
**Figure 5:** EPSPS interacting partners as well as its coexpression genes predicted by STRING. (A) Rice (B) *E. coli* (C) The key to the putative interacting partners for OsEPSPS gene is listed. (D) The key to the putative interacting partners of *E. coli* EPSPS gene is listed.
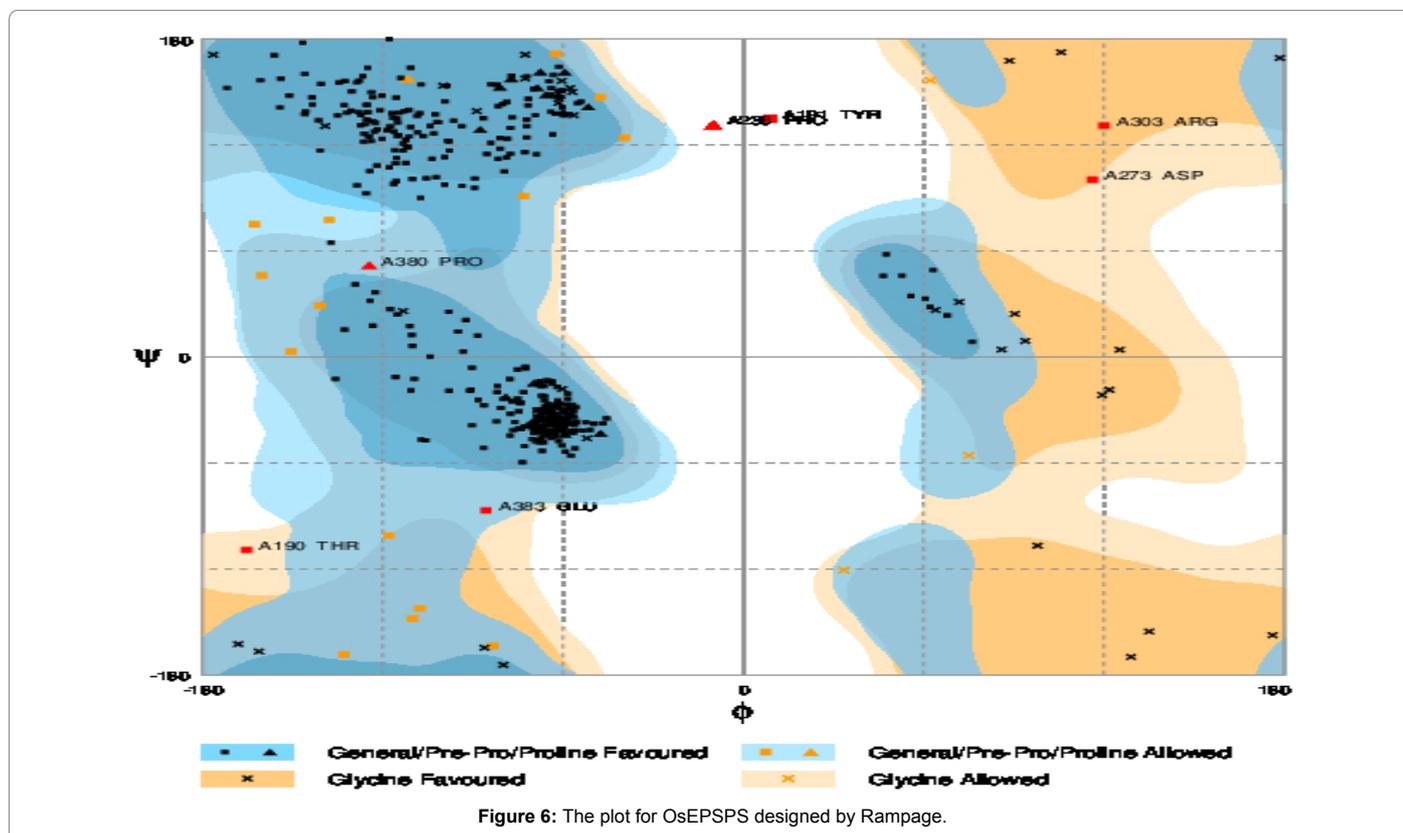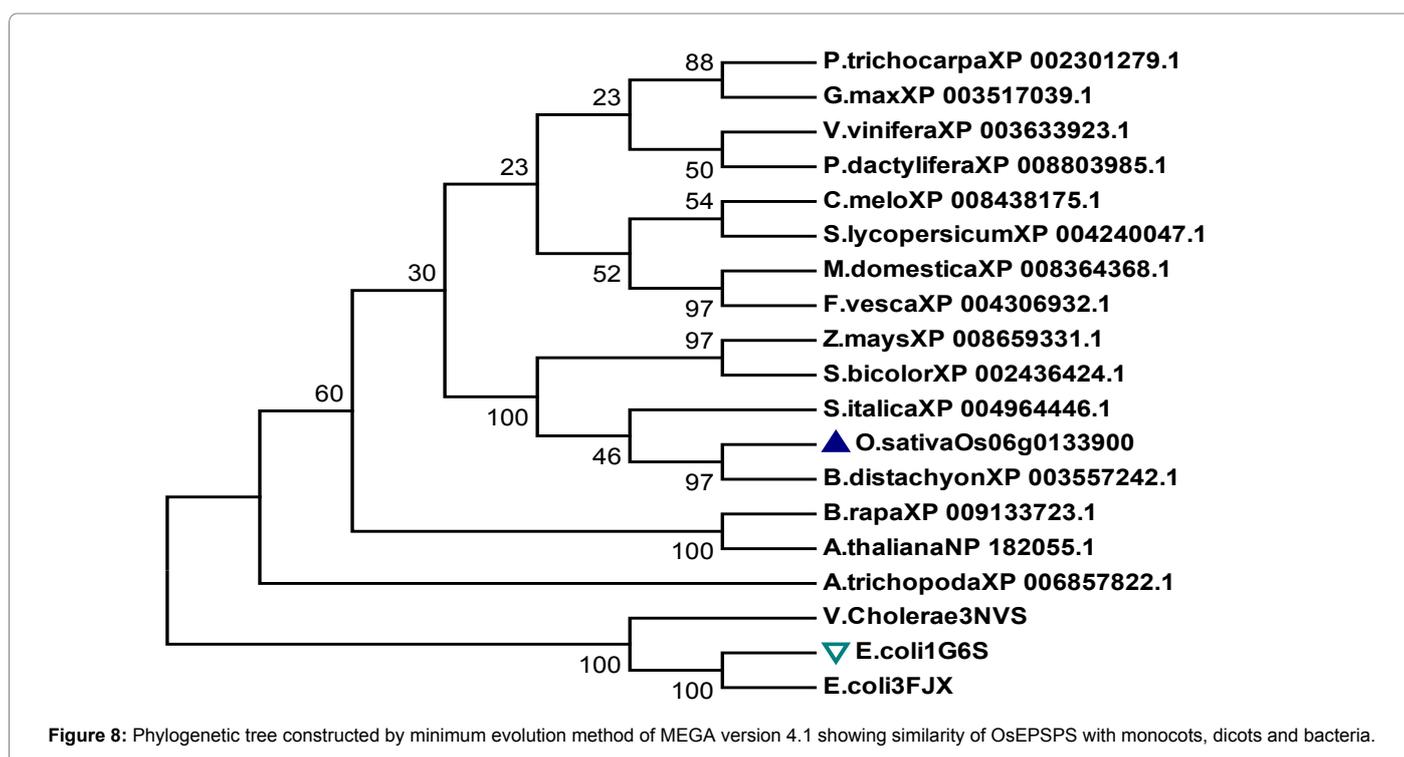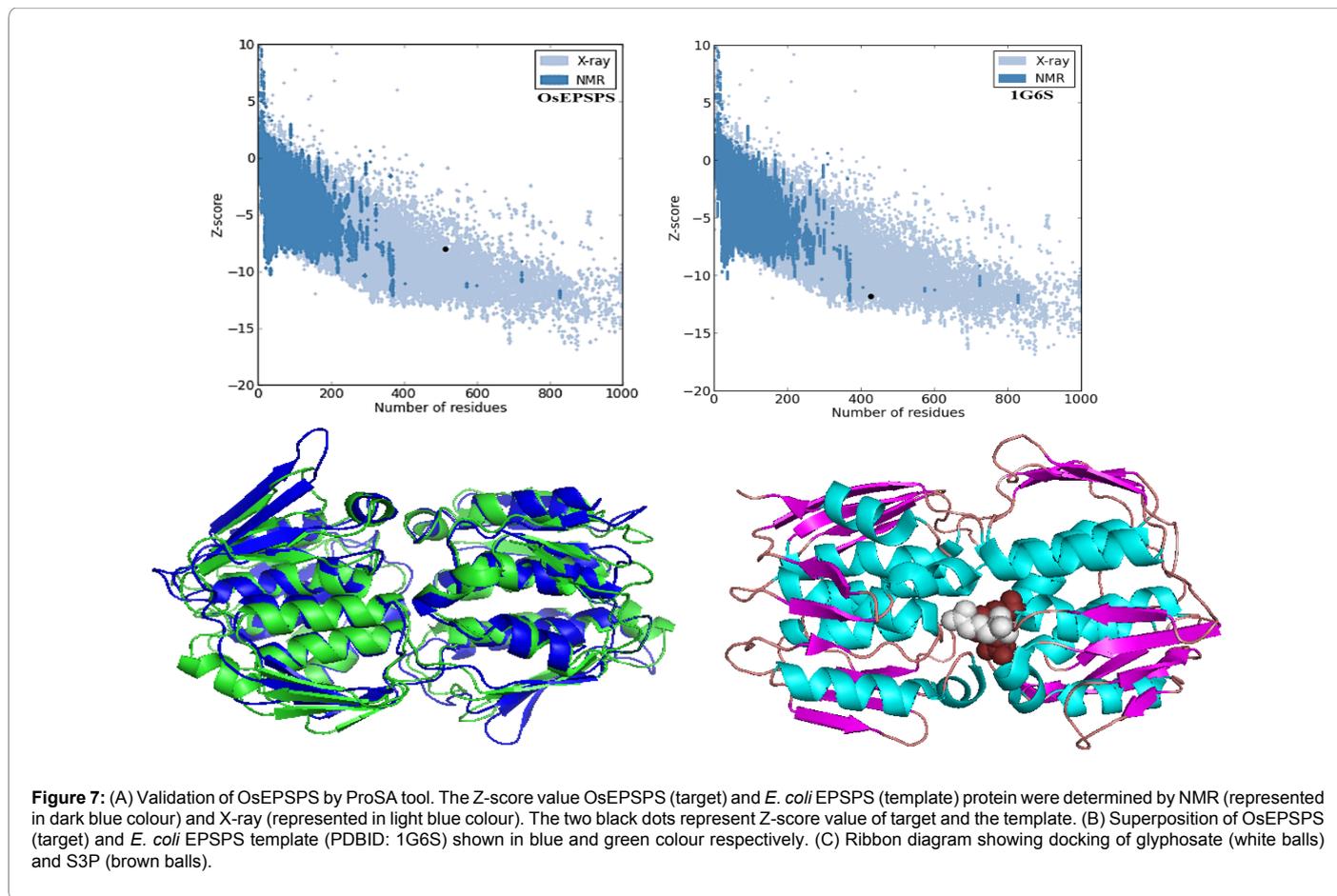


**Figure 6:** The plot for OsEPSPS designed by Rampage.

**Figure 7:** (A) Validation of OsEPSPS by ProSA tool. The Z-score value OsEPSPS (target) and *E. coli* EPSPS (template) protein were determined by NMR (represented in dark blue colour) and X-ray (represented in light blue colour). The two black dots represent Z-score value of target and the template. (B) Superposition of OsEPSPS (target) and *E. coli* EPSPS template (PDBID: 1G6S) shown in blue and green colour respectively. (C) Ribbon diagram showing docking of glyphosate (white balls) and S3P (brown balls).



**Figure 8:** Phylogenetic tree constructed by minimum evolution method of MEGA version 4.1 showing similarity of OsEPSPS with monocots, dicots and bacteria.

| Properties | OsEPSPS | 1G6S |
|---|---|---|
| Molecular weight | 54345.7 | 46095.7 |
| Theoretical pI | 8.04 | 5.37 |
| Number of amino acids | 515 | 427 |
| -R | 55 | 48 |
| +R | 57 | 38 |
| Aliphatic index | 93.42 | 94.66 |
| Grand average of hydropathicity (GRAVY) | 0.101 | -0.005 |
| Extinction coefficients ($M^{-1}$ $cm^{-1}$) | 34755 | 30745 |
| Instability index | 33.83 | 28.78 |
| CELLO predicted location | Combined | Combined |
| Predicted N-glycosylation sites | 188 NATY, 464 NITA | - |

**Table 3:** Physiochemical, structural and sequence properties, sub-cellular localizations and N-glycosylation sites of the EPSPS protein sequences.

| Ligand Name | Binding Residues |
|---|---|
| S3P | 94K 95S 99R 173T 249S 250S 251Q 277S 280Y 402D 429K |
| GPJ | 94K 170N 172G 202R 251Q 402D 430E 433R 474H 475R 500K |
| $PO_4$ | 168L 169G 170N 171A 196V 199M |

**Table 4:** Binding residues of different ligands of the OsEPSPS protein.

*sativa, Populus trichocarpa, Phoenix dactylifera, Setaria italica, Sorghum bicolor, Solanum lycopersicum, Vitis vinifera, Zea mays, E. coli and V. chloerae.* Many of these exhibited orthologous and paralogous relations with each other (Figure 8). However, *B. distachyon* showed highest sequence similarity to OSEPSPS. *Amborella trichopoda* is believed to be the most basal lineage in the clade of angiosperms. The results indicate that EPSPS protein gene family is strictly conserved and has evolved from bacteria.

## Acknowledgements

## Conflict of Interest

We declare that we have no conflict of interest.

## References

1. Zhou M, Xu H, Wei X, Ye Z, Wei L et al. (2006) Identification of a glyphosate resistant mutant of rice 5-enolpyruvylshikimate 3-phosphate synthase using a directed evolution strategy. Plant Physiol 140: 184-195.

2. Cao G, Liu Y, Zhang S, Yang X, Chen R et al. (2012) A novel 5-enolpyruvylshikimate-3-phosphate synthase shows high glyphosate tolerance in Escherichia coli and tobacco plants. PLoS ONE 7: 38718.

3. Bentley R (1990) The shikimate pathway--a metabolic tree with many branches. Crit. Rev. Biochem. Mol. Biol. 25: 307-383.

4. Walsh CT, Benson TE, Kim DH, Lees WJ (1996) The versatility of phosphoenolpyruvate and its vinyl ether products in biosynthesis. Chem. Biol. 3: 83-91.

5. Boocock MR, Coggins JR (1983) Kinetics of 5-enolpyruvylshikimate-3-phosphate synthase inhibition by glyphosate. FEBS Lett. 154: 127-133.

6. Steinrucken HC, Amrhein N (1984) 5-EnoZpyruvylshikimate-3-phosphate synthase of Klebsiella pneumonia 2 Inhibition by glyphosate N-(phosphonomethyl) glycine. Eur. J. Biochem. 143: 351-357.

7. Baerson SR, Rodriguez DJ, Tran M, Feng Y, Biest NA, et al. (2002) Glyphosate resistant goose grass. Identification of a mutation in the target enzyme 5-enolpyruvylshikimate-3-phosphate synthase. Plant Physiol 129: 1265-1275.

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

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

10. Coggins JR, Abell C, Evans LB, Frederickson M, Robinson DA, et al. (2003) Experiences with the shikimate-pathway enzymes as targets for rational drug design. Biochem Soc Trans 31: 548-552.

11. Westbrook J, Feng Z, Jain S, Bhat TN, Thanki N, et al (2002) The Protein Data Bank: unifying the archive. Nucleic Acids Res. 30: 245-248.

12. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389-3402

13. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680.

14. Thompson JD, Gibson TJ, Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX. Curr. Protoc. Bioinformatics Chapter 2, Unit 2.3.

15. Kawabata T (2003) MATRAS: a program for protein 3D structure comparison. Nucleic Acids Res. 31: 3367-3369.

16. Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analysing DNA and protein sequence motifs. Nucleic Acids Res 34: W369-W373.

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

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

19. Gill SC, Von Hippel PH (1989) Extinction coefficient. Anal Biochem 182: 319-328.

20. Guruprasad K, Reddy BVP, Pandit MW (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.

21. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, WilkinsMR, et al. (2005) Protein identification and analysis tools on the ExPASy server. In: Walker JM (ed.), The proteomics protocols handbook. Humana Press, Totowa 571-607.

22. Yu CS, Chen YC, Lu CH, Hwang JK (2006) Prediction of protein subcellular localization. Proteins 64: 643-651.

23. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, et al. (2013) STRING V9.1: Protein-protein interaction networks, with increased coverage and integration. Nuc Acid Res 41: 1.

24. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM (1992) Stereochemical quality of protein structure coordinates. Proteins 12: 345-364.

25. Eisenberg D, Luthy R, Bowie JU (1997) VERIFY3D: Assessment of protein models with three-dimensional profiles. Methods Enzymol 277: 396-404.

26. Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35: W407-W410.

27. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera-a visualization system for exploratory research and analysis. J. Comput. Chem 25: 1605-1612.

28. Maiti R, Domselaar GHV, Zhang H, Wishart DS (2004) SuperPose: a simple server for sophisticated structural superposition. Nucleic Acids Res 32: 590-594.

29. Liang J, Edelsbrunner H, Woodward C (1998) Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. Protein Sci 7: 1884-1897.

30. Laurie AT, Jackson RM (2005) Q-SiteFinder: an energy-based method for the prediction of protein ligand binding sites. Bioinformatics 21: 1908-1916.

31. Liang S, Zhang C, Liu S, Zhou Y (2006) Protein binding site prediction with an empirical scoring function. Nucleic Acids Res. 34: 3698-3707.

32. Homer RW, Swanson J, Jilek RJ, Hurst T, Clark RD (2008) SYBYL line notation (SLN): a single notation to represent chemical structures, queries, reactions, and virtual libraries. J. Chem. Inf. Model. 48: 2294-2307.

33. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4, Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Mol Biol Evol 24: 1596-1599.

34. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26: 283-291.

35. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, et al. (2006) Comparative protein structure modeling with Modeller. Curr Protoc Bioinformatics. Chapter 5, Unit 5.6.

36. Stallings WC, Abdel-Meguid SS, Lim LW, Shieh HS, Dayringer HE, et al. (1991) Structure and topological symmetry of the glyphosate target 5-enolpyruvylshikimate-3-phosphate synthase: a distinctive protein fold. Proc Natl Acad Sci USA 88: 5046-5050.

37. Gong Y, Liao Z, Chen M, Guo B, Jin H, et al. (2006) Characterization of 5-enolpyruvylshikimate 3-phosphate synthase gene from Camptotheca acuminate. Biol Plantarum 50: 542-550.

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

39. Filiz E, Koc I (2015) Genome-wide identification and comparative analysis of EPSPS (aroA) genes in different plant species. J Plant Biochem Biotechnol.

40. Liang L, Wei L, Yunlei H, Shuzhen P, Wei Z, et al. (2006) A novel RPMXR motif among class II 5-enolpyruvylshikimate-3-phosphate synthases is required for enzymatic activity and glyphosate resistance. J Biotech 144: 330-336.

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

42. Schwarz F, Aebi M (2011) Mechanisms and principles of N-linked protein glycosylation. Curr Opin Struct Biol 21: 576-582.

43. Schmid J, Amrhein N (1995) Molecular organization of the shikimate pathway in higher plants. Phytochem 39: 737-749.

44. Fiser A, Feig M, Brooks CL, Sali A (2002) Evolution and physics in comparative protein structure modeling. Acc Chem Res. 35: 413-421.

45. Shuttleworth WA, Pohl ME, Helms GL, Jakeman DL, Evans JNS (1999) Site-directed mutagenesis of putative active site residues of 5-enolpyruvylshikimate-3-phosphate synthase. Biochemistry 38: 296-302.