



Full Length Article

Computational Study of *Oryza sativa* Germin Like Protein 1 (OsRGLP1), from Genome Sequence to Protein Structure; Modeling and Interaction

Dure Shahwar, Syed Irtiza Hussain Shah Gardezi, Ghazala Kaukab Raja and Syed Muhammad Saqlan Naqvi*

University Institute of Biochemistry and Biotechnology, Pir Mehr Ali Shah Arid Agriculture University Rawalpindi, Pakistan

*For correspondence: saqlan@uaar.edu.pk

Abstract

Germin like proteins are components of a superfamily of glycosylated multimeric proteins of plant origin mainly exhibiting superoxide dismutase activity. Physiochemical characterization revealed that OsRGLP1 is a hydrophilic heat stable protein as predicted from protein low Instability Index (InI), high Aliphatic Index (AI) and Grand Average of Hydropathy (GRAVY) index. One N-linked glycosylation site in the N-terminal half of the OsRGLP1 protein was predicted using Ensembles of Support Vector Machine classifiers. InterProScan was used to predict the OsRGLP1 similarity to other germins and GLPs. Investigation of secondary structures revealed that random coils are highly abundant followed by extended β strand and alpha helix conformation, respectively. YASARA was employed to build 3-D structure of OsRGLP1 protein. To study molecular interaction and binding pattern of Manganese, a renowned molecular docking approach was applied. Functionally important residues and the involvement of histidine and glutamate residues in metal ion binding were determined by I-TASSER and INTREPID. Predicted amino acids were further evaluated by executing mutational analysis using HOPE web server which confirmed the importance of these residues in metal ion binding. Prediction of functionally important residues in OsRGLP1 may be exploited in wet lab for designing site directed mutagenesis experiments. These findings may further facilitate the understanding of catalytic mechanism and improvement of the enzymatic properties of the protein for commercial applications. © 2018 Friends Science Publishers

Keywords: Homology modeling; Binding site; Glycoprotein; Molecular dynamics

Introduction

The Germin and germin-like proteins represent a large protein family that comprised of more than twenty five members. Germins and GLPs are mostly glycosylated and normally stable in oligomeric forms. In contrast to germin proteins that are expressed only in monocots, members of germin-like proteins are found in whole plant kingdom including gymnosperm as well as angiosperms. Expression of germin-like proteins was observed in different parts of the host plant and at different developmental stages. Most of the germins are thought to reside in the extracellular matrix in a non-covalently attached manner (Membre *et al.*, 2000; Dunwell *et al.*, 2008).

Germin and GLPs are members of cupin superfamily having a conserved beta-barrel core and difference in the region of active site endow GLPs with varieties of non enzymatic and enzymatic functions (Patnaik and Khurana, 2001; Zimmermann *et al.*, 2006). The exact role of GLPs in plant metabolism is poorly understood. Increasing evidence showed that GLPs may be acting as enzymes (superoxide dismutase) opening new vista in commercial application where the particular interest lies on their significant resistance to denaturation. Germin-like proteins may

provide first line of defense against biotic and abiotic stresses, by dismutating superoxide anions into hydrogen peroxide and molecular oxygen (Dunwell, 1998).

Superoxide dismutases (SOD) are metalloenzyme, which convert superoxide radical into hydrogen peroxide and molecular oxygen resulting in protection of plants from the damage cause by reactive oxygen species (ROS). Metal ion is found to be present in the active site of SOD. On the basis of metal ion, SOD are classified into three main types i.e. the Cu/Zn SOD (mostly found in eukaryotes), Fe-SODs (found both in prokaryotes and eukaryotes i.e. higher plants) and Mn-SODs or Fe/ Mn-SODs (found both in prokaryotes and eukaryotes) (Alscher *et al.*, 2002).

Active site of any enzyme is a location in protein that binds with ligand and where the chemical reaction takes place. As in enzyme catalyzed reaction just a small fraction of amino acid residues participate, so physical, chemical properties and location of these residues in active site decides the nature of catalytic reaction. Therefore, identification of catalytic residues is an important step for understanding the mechanism of enzyme catalyzed reactions (Chou and Cai, 2004).

Protein sequences of many germin-like proteins, is

available in protein sequence databases but information about the 3-D structure and function is limited. Determination of 3-D structure and function of these proteins by using expensive and laborious experimental techniques is time consuming hence to facilitate information gathering, bioinformatics tools have been exploited in the present study to analyze one of the GLPs, namely, *OsRGLP1*.

In current work, we investigated physiochemical properties of *OsRGLP1*, secondary and tertiary structure, protein domain and perform function analysis. These findings support the hypothesis that *OsRGLP1* is like other germin proteins in having characteristic β -barrel structure and by the involvement of similar residues at the β -barrel fold, in constituting its active site.

Materials and Methods

Computational Analyses of *OsRGLP1* Physiochemical Properties

For computing basic physiochemical properties of *OsRGLP1* based on instability index, theoretical isoelectric point (pI), aliphatic index, molecular weight and grand average of hydropathicity (GRAVY), ExPASy-ProtParam tool Proteomic server was used (Gasteiger *et al.*, 2003). For C-, N-, and O-linked glycosylation prediction tasks, Ensembles of Support Vector Machine classifiers offer was employed as it offers consistent and correct approach for automated identification of putative glycosylation sites in glycoprotein sequence information (Caragea *et al.*, 2007). InterProScan tool was used to locate conserved domains in *OsRGLP1* (Quevillon *et al.*, 2005). For the prediction of secondary structure in *OsRGLP1*, two important computational algorithms were exploited, namely, (i) Self Optimized Prediction Method with Alignment (SOPMA) and (ii) PSIPRED Protein Structure Prediction server (Geourjon and Deleage, 1995; McGuffin *et al.*, 2000).

Prediction of *OsRGLP1* Three Dimensional Structure

The homology model of *OsRGLP1* was built using YASARA (Yet Another Scientific Artificial Reality Application) software (Elmar *et al.*, 2002). Possible templates (Crystal structure of oxalate oxidase PDB id: 2ET1-A, 2ET7-A and 2ETE-A) were identified by YASARA using 3 PSI-BLAST iterations to mine a position specific scoring matrix (PSSM) from UniRef90, followed by searching the PDB for a similarities (i.e., matches with an E-value below the cutoff (0.5) of homology modeling). For each of the templates selected, models were built. Either a single model if the alignment was certain or a number of alternative models if the alignment was ambiguous. Finally, YASARA tried to combine the best parts of all the models to obtain a hybrid model, hoping to increase the accuracy beyond each of the contributors.

Prediction of Functionally Important Residues in *OsRGLP1*

The Inference of Evolutionary conserved residues (INTREPID) tool was used to predict functionally important residues in *OsRGLP1* (Sankararaman *et al.*, 2009) and I-TASSER (Zhang, 2008) web servers. Structurally and functionally important residues predicted by INTREPID and I-TASSER, were replaced with glycine residue. Glycine was selected because it is a simple amino acid without any side chain. Therefore, the effect of mutation on SOD activity was only due to removal of above mentioned four residues not due to the presence of incompatible side chain of another residue. Mutations were analyzed using HOPE, a self-supported program that perform functional and structural analysis of mutations at single position. It gathers data from diverse sources including WHAT IF calculations based on the three dimensional coordinates of the protein, UniProt database sequence annotations and predictions of DAS services. Data was accumulated in a database and further used for mutant proteins structure and function analysis (Venselaar *et al.*, 2010).

Molecular Docking Studies

Molecular structure of glycolate, Mn and oxalic Acid was drawn using ChemOffice 8.0 Ultra, while MOPAC2009 and RMI semi-empirical method were used to minimize the energy. In order to perform molecular/flexible docking, receptor and ligands; Glycolate, Mn & Oxalic Acid were used in YASARA. Grid box was set for each model on total receptor, with measuring X, Y, Z co-ordinates as 126°, 90° and 80°, respectively. While setting the grid, various docking experiments were performed, including the peripheral helical motifs. A hybrid of Lamarckian and Genetic Algorithm is used by YASARA with 150 rounds of docking for each, with maximum energy evaluations equal to 2,500,000, RMSD cut off of 2.0 Å°, cross-over and mutation rates (recombination) equal to 0.8 and 0.02 respectively. The docking was performed keeping the ligand flexible and the receptor as rigid molecule, with exclusion of non-interacting water molecules. Only eleven water molecules found in binding pocket were kept due to their role in interaction.

Results

Physiochemical Characterization of *OsRGLP1*

The table 1 summarizes the ProtParam analysis of physiochemical properties of mature *OsRGLP1* (without signal peptide). After removing signal peptide of approximately 19 amino acids, mature *OsRGLP1* is comprised of about 205 amino acids having a molecular mass of 22 kD and isoelectric point of 6.3. Both the above mentioned information can be utilized in PAGE and

isoelectric focusing. It can be predicted from protein low Instability Index (InI) and high Aliphatic Index (AI) values that *OsRGLP1* is a heat stable protein. Grand average of hydropathicity (GRAVY) index indicates its higher surface accessibility to interact with water, supporting the assumption about its localization in extracellular matrix.

The information obtained from Ensembles of Support Vector Machine classifiers revealed the presence of one N-linked glycosylation site in the N-terminal half of the of *OsRGLP1* protein (Fig. 1). InterProScan predicted the similarity of *OsRGLP1* to other germins and GLPs in having two main features of cupin superfamily i.e. a cupin domain and metal ion binding site (Fig. 2) (Dunwell *et al.*, 2004).

Prediction of Secondary Structures in *OsRGLP1*

Self Optimized Prediction Method with Alignment (SOPMA) (Geourjon and Deleage, 1995) and (ii) PSIPRED Protein Structure Prediction server (McGuffin *et al.*, 2000) were used for *OsRGLP1* secondary structure prediction. Among all predicted secondary structures, random coils are highly abundant. The remaining secondary structure is represented by extended β strand and alpha helix conformation, respectively. The Secondary structure prediction was performed by using default parameters (similarity threshold: 8, window width 17) except number of conformational states that was chosen as 3 (random coil, extended strand and alpha helix) (Fig. 3).

Prediction of *OsRGLP1* Model

YASARA was utilized to build three dimensional model of *OsRGLP1*. After comparing the scores of all models, the best scoring model based on template with PDB id: 2ET1-A was selected as the final model (Fig. 4).

Identification of Functionally key Residues

Prediction of active site residues: I-TASSER and INTREPID predicts functional and catalytic residues with high precision using methods based on amino acid sequence (Fig. 4). These selection criteria for these residues was conservation score cutoff value of greater than 1.573. The analysis of the modeled *OsRGLP1* structure predicted essentially the same residues at the β -barrel fold involved in metal ion binding. It was also proved in a study conducted by Banerjee and Maiti (2010) that the active site of GLPs is buried inside nine β -strands of Jelly roll β -barrel structure.

Molecular Docking and Interaction Study

In case of ligands, non-polar hydrogen atoms were combined and Gasteiger partial charges were selected. All torsions for ligands were allowed to rotate during docking procedure. The grid maps were produced using AutoGrid program in YASARA. Each grid was positioned at the midpoint of the corresponding receptor structure.

Table 1: ProtParam analysis of physiochemical properties of *OsRGLP1*

Parameters	<i>OsRGLP1</i> without signal peptide region
Sequence length	205
Molecular weight	22261.4
Theoretical isoelectric point	6.30
Instability index	28.91
Aliphatic index	88.00
GRAVY	-0.001

```
>osRGLP1
MASSSFLLLATLLAMASWQGMASDPSPLQDFCVADMHSPV
LVNGFACLNPKDVENADHFFKAAMLDTPRKTNKVGSNVTLI
-----N-----

NVMQIPGLNTLGISARIDYAPLGQNPPHPTHPRATEILTVLEGT
LYVGFVTSNPDNKFVSKVLNKGDFVFPVGLIHFFQF
-----

NPNPYKPAVAIAALSSQNPGAITIANAVFGSKPPISDDVLAKAF
QVEKGTIDWLQAQFWENNHY
-----
```

Fig. 1: EnsembleGly server results showing the predicted N-linked glycosylation site (in red color) in *OsRGLP1* sequence



Fig. 2: InterProScan server results, showing the presence of cupin domain (in blue color) and manganese binding site (in yellow color) in *OsRGLP1*

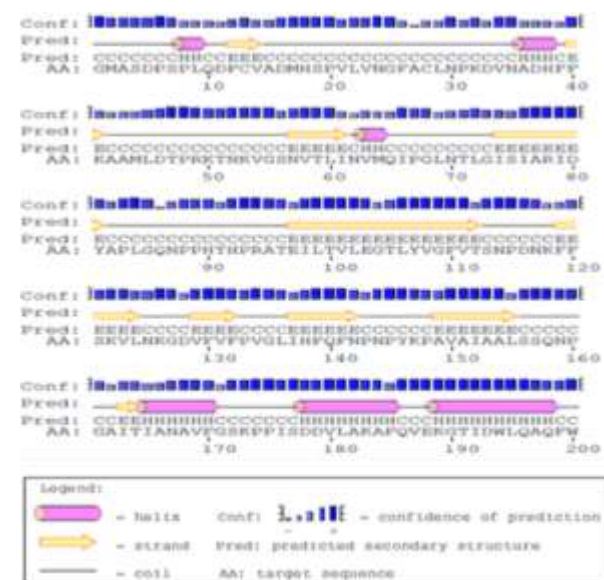


Fig. 3: Secondary structures features of *OsRGLP1*. Secondary structures predicted using PSIPRED showed high rate of occurrence of beta strand in comparison to alpha helix and coil

The grid dimensions were $120 \times 120 \times 120 \text{ \AA}^3$ with points separated by 0.375 \AA . Random torsions, orientations and starting positions were employed for all the ligands. Default values of the quaternion, translation and torsion steps were taken as indicated in YASARA. The Lamarckian genetic algorithm method was employed for minimization using default parameters (Morris *et al.*, 1998). For rigid and flexible ligand, standard docking protocol comprised of 100 runs, starting from a population of 150 individuals that were randomly placed with 2.5×10^6 energy evaluations, a maximum number of 27000 iterations, and 0.02 mutation rate, 0.80 crossover rate and an elitism value of 1. Docked results were subjected to cluster analysis using an RMS tolerance of 1.0 \AA . The clusters were ranked by the lowest energy representative of each cluster. Docking results were later analyzed and interaction was found to be in deep narrow groove (Fig. 5).

After completion of docking runs (100 for each ligand against *OsRGLP1*), specified orientations and resulted conformations were analyzed using (protein visualization and analysis tool) to know interacting residues details and binding pattern.

It was found that all ligands (Mn^{+2} , Glycolate and Oxalic Acid) bounded to same deep narrow groove (Fig. 5) surrounded by polar, acidic and hydrophobic residues present in C-terminal of *OsRGLP1*. Details of interaction and conformations are shown in Fig. 5. Hydrophobic interaction was calculated through binding free energy and inhibitory constant value. It was shown that Glycolate had binding free of -4.68 Kcal/mol and inhibitory constant of 37 pico-Molar, whereas Oxalic acid bounded in similar way with binding free energy of -4.5 Kcal/mol having inhibitory constant of 205 pico-Molar. Mn^{+2} showed weak interaction as it had binding free energy of -3.67 Kcal/mol .

Mutation Analysis

Amino acids identified to be important by I-TASSER and INTREPID were further evaluated by executing mutational analysis using HOPE web server. For mutational analysis, histidine residues at 109, 111 and 157 positions and glutamate at 116 position were replaced with Glycine. Glycine was selected because it is a simple amino acid without any side chain. Therefore, the effect of mutation on SOD activity was only due to removal of above mentioned four residues not due to the presence of incompatible side chain of another residue. Results of mutational analysis are summarized in Table 2. Computational analysis of mutants predicted complete loss of SOD activity of mutant proteins. In the 3D-structure it can be observed that the wild-type seemed to be involved in a metal-ion contact and new mutant residue disturbed the interaction with the manganese metal-ion: Mn^{+2} (Fig. 6).

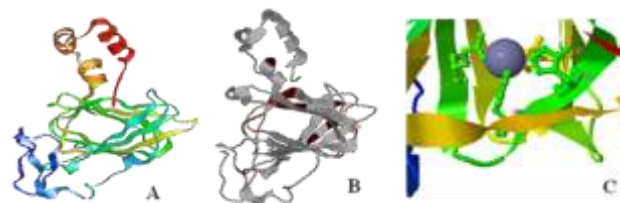


Fig. 4: Identification of functionally key residues. A. Ribbon presentation of predicted three dimensional structure of *OsRGLP1* without signal peptide. Residues adjacent to amino terminus are colored blue and those adjacent to Carboxyl terminus are colored red. Residues between N- and C-terminus are colored across the visible spectrum. Prediction of Active Site Residues B. using INTREPID. Top ranked residues are shown in red color on three dimensional structure of *OsRGLP1*. C. I-TASSER predicts 4 top-ranked *OsRGLP1* residues three histidine and one glutamate involved in metal ion binding i.e. manganese, shown as dark blue sphere

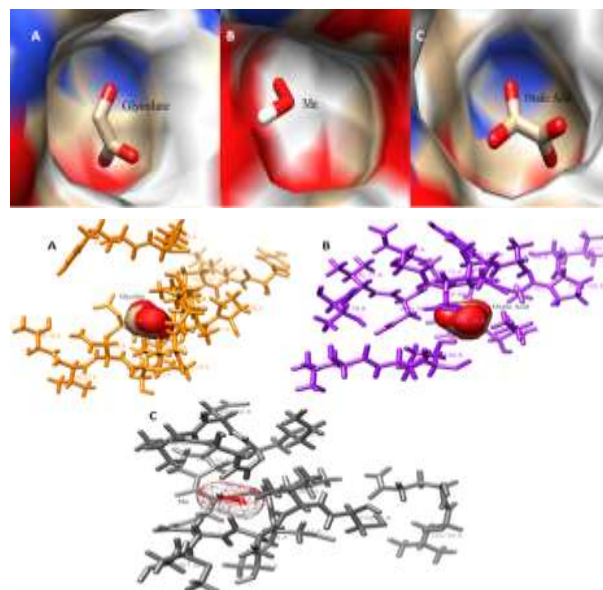


Fig. 5: Molecular Docking and Interaction study. Shows the hydrophobic interaction of glycolate, Mn, and oxalic acid in deep narrow groove of *OsRGLP1*

Discussion

In silico analysis of *OsRGLP1* illustrates a number of significant physiochemical, structural and functional properties of *OsRGLP1*. Analysis of its Physiochemical properties shows that *OsRGLP1* is a hydrophilic protein that correlates with the predicted location of germin-like proteins as they are believed to reside on the extracellular matrix. This is further confirmed by the presence of a signal peptide in GLPs that can direct the protein to the extracellular matrix (Berna and Bernier, 1997; Patnaik and Khurana, 2001).

Table 2: Mutational analysis of ligand binding residues of *OsRGLP1*

Mutation No.	Wild type residue	Residue position	Mutant residue	Effect of mutations
1	Glutamate	116	Glycine	Mutations at 116,109,157, and 111 positions directly or indirectly have profound effect on proteins' interaction with metal ion e.g., manganese and disturb its correct folding.
2	Histidine	109		
3	Histidine	157		
4	Histidine	111		

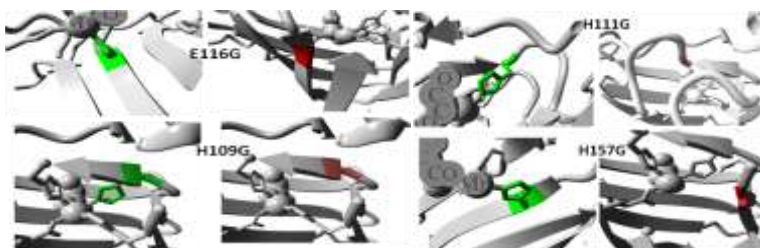


Fig. 6: Mutation analysis using HOPE web server. Original wild-type residues (glutamate and histidine) was replaced with Glycine residues. In the 3D-structure it can be seen that the wild-type residue was found to involve in a metal-ion contact and new mutant residue don't show any interaction with the metal ion. Wild type residue is shown in green color is involved in binding with metal ion that is represented as grey circles. Mutant (glycine) residue shown in red color, have no contact with metal ion

Values of its Instability index and aliphatic index predict that *OsRGLP1* is a stable protein and shows strong resistance to heat denaturation. Heat resistance of germin-like proteins was proved in various studies (Tabuchi *et al.*, 2003; Gucciardo *et al.*, 2007). Yasmin *et al.* (2015) also observed heat resistant SOD activity in transgenic tobacco plants expressing rice *OsRGLP1* protein. The information obtained from Ensembles of Support Vector Machine classifiers revealed the presence of one N-linked glycosylation site in the N-terminal half of the of *OsRGLP1* protein. Glycosylation plays essential role in secretion, solubilization, activation, folding, signaling and protein binding in various proteins (Lerouge *et al.*, 1998). We are still lacking in the information about the exact function of GLPs glycosylation but conservation of this site both in monocot and dicot species is observed that reflects the essential role of glycosylation in GLPs' function (Carter and Thornburg, 1999). Further studies are needed to find the nature and exact role of glycosylation in GLPs. Results of InterProScan server show cupin domain, a conserved barrel domain of the 'cupin' superfamily and manganese ion binding site in *OsRGLP1*. Using YASARA software three-dimensional structure of *OsRGLP1* model is also predicted. Predicted model show the resemblance of *OsRGLP1* to germin protein, in having characteristic β -barrel structure. Analysis of the functionally important residues in *OsRGLP1* structure predicted one Glutamate residue at 116 position and three Histidine residues at 109, 111 and 157 positions at the β -barrel fold putatively involved in ligand binding. It was also proved in a study conducted by Banerjee and Maiti (2010) that the active site of GLPs is buried inside nine β -strands of jelly roll β -barrel structure. Mutational analysis confirmed the significance of these residues in metal ion binding. Further wet lab analysis is required to confirm

these findings and these point mutations might further be explored by designing specific primers. Computational analysis of *OsRGLP1*, provided a key insight of its primary and tertiary structure, revealing its similarity to germin in structure. These results also raised a question that In spite of having structural similarity to germin, GLPs possess different enzymatic activity that need to be answered by further research and in silico analysis.

Conclusion

The present study can be helpful for protein purification for example predicted isoelectric point of *OsRGLP1* can be helpful in protein separation by isoelectric focusing on a polyacrylamide gel and helps in proper buffer selection. Further computational analysis of key residues responsible for constituting active site may provide insight of diverse enzymatic activities in Cupin Superfamily.

Acknowledgements

This work was supported by Higher Education Commission (HEC), Islamabad, Pakistan

References

- Alscher, R.G., G. Erturk and L.S. Heath, 2002. Role of superoxide dismutases (SODs) in controlling oxidative stress in plants. *J. Exp. Bot.*, 53: 1331–1341
- Banerjee, J. and M.K. Maiti, 2010. Functional role of rice germin-like protein1 in regulation of plant height and disease resistance. *Biochem. Biophys. Res. Commun.*, 394: 178–83
- Berna, A. and F. Bernier, 1997. Regulated expression of a wheat germin gene in tobacco: oxalate oxidase activity and apoplastic localization of the heterologous protein. *Plant Mol. Biol.*, 33: 417–429

- Caragea, C., J. Sinapov, A. Silvescu, D. Dobbs and V. Honavar, 2007. Glycosylation site prediction using ensembles of support vector machine classifiers. *B.M.C. Bioinform.*, 8: 438
- Carter, C. and R.W. Thorburg, 1999. Germin-like proteins: Structure, phylogeny, and function. *J. Plant Biol.*, 42: 97–108
- Chou, K.C. and Y.D. Cai, 2004. A novel approach to predict active sites of enzyme molecules. *Proteins*, 1: 77–82
- Dunwell, J.M., 1998. Cupins: a new superfamily of functionally diverse proteins that include germins and plant storage proteins. *Biotechnol. Genet. Eng. Rev.*, 15: 1–32
- Dunwell, J.M., A. Purvis and S. Khuri, 2004. Cupins: the most diverse protein superfamily? *Photochemistry*, 65: 7–17
- Dunwell, J.M., J.G. Gibbings, T. Mahmood and S.M.S. Naqvi, 2008. Germin and germin-like proteins: Evolution, structure, and function. *Crit. Rev. Plant Sci.*, 27: 342–375
- Krieger, E., G. Koraimann and G. Vriend, 2002. Increasing the precision of comparative models with YASARANOVA a Self-Parameterizing force field. *Proteins: Struct. Funct. Genet.*, 47: 393–402
- Gasteiger, E., A. Gattiker, C. Hoogland, I. Ivanyi, R.D. Appel and A. Bairoch, 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.*, 31: 3784–3788
- Geourjon, C. and G. Deléage, 1995. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biosci.*, 11: 681–684
- Gucciardo, S., J.P. Wisniewski, N.J. Brewin and S. Bornemann, 2007. A germin-like protein with superoxide dismutase activity in pea nodules with high protein sequence identity to a putative rhicadhesin receptor. *J. Exp. Bot.*, 58: 1161–1171
- Lerouge, P., M.C. Macheteau, C. Rayon, A.C. Fischette, V. Gomord and L. Fayel, 1998. N-glycoprotein biosynthesis in plants: recent developments and future trends. *Plant Mol. Biol.*, 38: 31–48
- McGuffin, L.J., K. Bryson and D.T. Jones, 2000. The PSIPRED protein structure prediction server. *Bioinformatics*, 16: 404–405
- Membre, N., F. Bernier, D. Staiger and A. Berna, 2000. *Arabidopsis thaliana* germin-like protein: common and specific features point to a variety of functions. *Planta*, 211: 345–354
- Morris, G.M., D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew and A.J. Olson, 1998. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.*, 19:1639–1662.
- Patnaik, D. and P. Khurana, 2001. Germins and germin-like proteins: an overview. *Ind. J. Exp. Biol.*, 39: 191–200
- Quevillon, E., V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler and R. Lopez, 2005. InterProScan: protein domains identifier. *Nucleic Acids Res.*, 33: 116–120
- Sankararaman, S., B. Kolaczowski and K. Sjölander, 2009. INTREPID: a web server for prediction of functionally important residues by evolutionary analysis. *Nucleic Acids Res.*, 7: 390–395
- Tabuchi, T., T. Kumon, T. Azuma, T. Nanmori and T. Yasuda, 2003. The expression of a germin like protein with superoxide dismutase activity in the halophytes *Atriplex lentiformis* is differentially regulated by wounding and abscisic acid. *Physiol. Plant.*, 118: 523–531
- Venselaar, H., T.A. Te Beek, R.K. Kuipers, M.L. Hekkelman and G. Vriend, 2010. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *B.M.C. Bioinform.*, 11: 548
- Yasmin, T., A. Mumtaz, T. Mahmood, M.Z. Hyder and S.M.S. Naqvi, 2015. A germin-like protein gene of rice increased superoxide dismutase activity in transformed tobacco. *Biol. Plant.*, 59: 456–462
- Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. *B.M.C. Bioinform.*, 9: 40
- Zimmermann, G., H. Baumlein, H.P. Mock, A. Himmelbach and P. Schweizer, 2006. The multigene family encoding germin-like proteins of barley. Regulation and function in basal host resistance. *Plant Physiol.*, 142: 181–192

(Received 21 February 2017; Accepted 28 August 2017)