In silico analysis of the germinlike protein multigene family members of tomato with predicted oxalate oxidase activity

Tirthartha Chattopadhyay

Department of Plant Breeding and Genetics, Bihar Agricultural University, Sabour, Bhagalpur, Bihar, India-813210

Corresponding author: tirthartha@gmail.com

Paper No. 255 Received: 3 July, 2014 Accepted: 15 October, 2014 Published: 17 December, 2014

Abstract

Germins and germin like proteins (GLPs) have been reported as plant glycoproteins belonging to the ‘cupin’ superfamily. They have been documented to possess enzymatic activities leading to the generation of \( \text{H}_2\text{O}_2 \), a reactive oxygen species (ROS). Hence, members of the GLP family have been proposed to play major role in plant disease resistance through ROS-mediated signaling. Interestingly, the functional characterization of GLP(s) in terms of their suitability as a key player in plant disease resistance has remained under-explored in case of tomato (Solanumlycopersicum L.). In the present study, 15 tomato GLPs, predicted to have oxalate oxidase activity, have been investigated in silico. Deduced multiple amino acid sequence alignment-based clustering of these proteins was carried out to classify them into 3 sub-families. All the proteins were found to contain the conserved amino acid stretches, representing the BoxA, B and C, and an inter-motif region of variable length. It was observed through homology modeling and structural alignment that the active sites of all except 3 of these proteins have Mn\(^{2+}\) bound at the active site involving the three conserved histidine and one conserved glutamate residue(s). The active site architecture was analyzed with a comparative view in order to examine the metal binding capacity of tomato GLPs. Thus the present work makes a platform for further genetic, molecular biological and functional genomics studies in the field of tomato GLPs, the possible key players for conferring biotic and/or abiotic stress tolerance, in future.

Highlights

• In tomato, 15 genes encoding germin like proteins (GLPs) with predicted oxalate oxidase activity were found to be distributed in 5 different chromosomes.
• The 15 tomato GLPs were found to have structural distinctness with a basic conserved ‘jelly roll’ structure.
• Three out of the 15 proteins were predicted not to bind Mn\(^{2+}\).
• The distinct active site architecture of the tomato GLPs might explain the differential Mn\(^{2+}\) binding.

Keywords: Active site architecture, biotic stress, homology modeling, Mn\(^{2+}\) binding, point mutation.
properties, this class of proteins has been considered quite crucial for plants' resistance to several biotic stresses. Several GLPs have been demonstrated to have oxalate oxidase (OXO) (Dumas et al., 1993; Lane et al., 1993) or superoxide dismutase (SOD) (Yamahara et al., 1999; Carter and Thornburg 2000; Tabuchi et al., 2003; Christensen et al., 2004) activities. The OXO and/or SOD activity-generated \( \text{H}_2\text{O}_2 \) has been reported to be involved in disease resistance of plants (Lamb and Dixon, 1997) through reaching a toxic level that can hinder the growth of the microbes (Peng and Ku’c, 1992). At the same time, the elevated level of \( \text{H}_2\text{O}_2 \), a reactive oxygen species (ROS), has been documented to cause structural modification of the plant cell wall (Bolwell et al., 1995), induce lipid peroxide, salicylic acid, ethylene synthesis and finally to activate the programmed cell death (PCD) in plant cells (van Breusegem et al., 2001). Furthermore, \( \text{H}_2\text{O}_2 \) is supposed to play important role in signal transduction leading to induction of pathogenesis-related proteins, polyphenol oxidase and hypersensitive response.

The GLPs have been well studied and characterized in different plant species. Thirty two GLP genes have been identified from *Arabidopsis* (Carter et al., 1998; Carter and Thornburg 1999), 14 have been documented from *Hordeum vulgare* (Wu et al., 2000; Druka et al., 2002; Zimmermann et al., 2006), and around 8 different GLP genes have been reported to be expressed in rice (Membré and Bernier, 1998). The GLPs have also been studied in grape vine (Godfrey et al., 2007), pea (Gucciardo et al., 2007), wheat (Segarra et al., 2003) and many other plant systems. Several germins and GLPs have been described at functional level through transgenic research. Transgenic soybean (Donaldson et al., 2001) poplar (Liang et al., 2001) and sunflower (Hu et al., 2003) plants expressing the wheat germ have been reported to have enhanced tolerance to pathogen attack. Similarly, transgenic maize plants over-expressing wheat germ have been documented to exhibit enhanced resistance to corn borer (Ramputh et al., 2002). Recently, it has been reported that silencing the germin like protein in *Nicotiana attenuata* results in improvement of the performance of herbivores (Lou and Baldwin, 2006). In case of rice, the GLP multigene family has been characterized to act as a complex quantitative trait loci (QTL) for disease resistance (Manosalva et al., 2009). Among the rice GLPs, OsGLP1 has been documented to play a role in determining plant architecture (plant height) as well as disease resistance (Banerjee and Maiti, 2010). The OsGLP1 has been found to possess superoxide dismutase (SOD) activity and to be involved in cell wall reinforcement in transgenic tobacco plants expressing the same (Banerjee et al., 2010). In case of *Brassica*, the GLP family members has been reported to be involved in initiating oxidative burst to impede pathogenesis of the pathogen *Sclerotinia sclerotiorum* (Rietz et al., 2012). All these observations make GLPs an interesting class of protein to work on.

Among the major vegetable crops, tomato (*Solanum lycopersicum* L.) is one with a global importance. Recent decoding of the genome sequence in tomato has broadened the scope of plant molecular biological research in this important crop plant. Taking the help of the decoded genome, the present study documents the in silico characterization of a multigene family of tomato encoding GLPs with predicted oxalate oxidase activity. Through homology-based structural analysis, the metal binding active sites of these proteins are analyzed to explain the predicted differential metal binding capacity. Along with that, point mutations leading to the same are identified. As a detailed bioinformatics analysis is rendered as a pre-requisite for any molecular biological research, this attempt might pave the way for further research on tomato GLPs in near future.

**Materials and Methods**

Retrieving the sequence information of the 15 tomato GLPs with predicted oxalate oxidase activity was done using the SolGenomics database (http://solgenomics.net/). Annotation of them was done as SIGLPA...SIGLPO in this study. The accession numbers of them are SIGLPA (Solyc01g088280.1), SIGLPB (Solyc01g088290.1),
In silico analysis of the germin-like protein

SlGLPC (Solyc01g088300.2), SlGLPD (Solyc01g102900.2), SlGLPF (Solyc01g102910.2), SlGLPG (Solyc03g123410.1), SlGLPH (Solyc07g041720.1), SlGLPI (Solyc07g065330.2), SlGLPJ (Solyc09g089990.2), SlGLPK (Solyc09g090000.1), SlGLPL (Solyc09g090010.2), SlGLPM (Solyc09g090020.2), SlGLPN (Solyc09g090040.2) and SlGLPO (Solyc11g068600.1). The GenBank accession numbers of the barley (Hordeum vulgare) GLPs, used during cluster analysis are HvGER1a (ABG46232), HvGER2a (ABG46233), HvGER3a (ABG46234), HvGER4c (ABG46235), HvGER5a (ABG46237) and HvGER6a (ABG46238).

Multiple amino acid sequence alignment and analysis were done through the Clustal Omega web server (https://www.ebi.ac.uk/Tools/msa/clustalo/) and pictorial representation was done using the EsPript web-server (http://espript.ibcp.fr). Analysis of different parameters associated with the protein sequences were carried out using the ExPasy web-resources (www.expasy.org). Performance of homology-based structural modeling was done in the SWISS MODEL web-server (http://swissmodel.expasy.org/) and pictorial representation was done through PyMol (www.pymol.org).

Table 1. Properties of the 15 tomato GLP genes and their deduced amino acid sequences

<table>
<thead>
<tr>
<th>Name</th>
<th>Located on (chromosome number)</th>
<th>Gene length (bp)</th>
<th>No. of introns</th>
<th>CDS* length (bp)</th>
<th>Predicted molecular weight of polypeptide (kDa)</th>
<th>Predicted theoretical pI of polypeptide</th>
<th>Predicted signal peptide cleavage site (between amino acid residue numbers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIGLPA</td>
<td>1</td>
<td>688</td>
<td>1</td>
<td>639</td>
<td>23.95</td>
<td>9.21</td>
<td>-</td>
</tr>
<tr>
<td>SIGLPB</td>
<td>1</td>
<td>546</td>
<td>1</td>
<td>405</td>
<td>14.86</td>
<td>8.49</td>
<td>-</td>
</tr>
<tr>
<td>SIGLPC</td>
<td>1</td>
<td>789</td>
<td>1</td>
<td>678</td>
<td>24.55</td>
<td>7.79</td>
<td>22-24</td>
</tr>
<tr>
<td>SIGLPD</td>
<td>1</td>
<td>774</td>
<td>1</td>
<td>687</td>
<td>24.52</td>
<td>5.61</td>
<td>22-23</td>
</tr>
<tr>
<td>SIGLPE</td>
<td>1</td>
<td>770</td>
<td>1</td>
<td>687</td>
<td>24.39</td>
<td>7.80</td>
<td>22-23</td>
</tr>
<tr>
<td>SIGLPF</td>
<td>1</td>
<td>793</td>
<td>1</td>
<td>687</td>
<td>24.63</td>
<td>6.16</td>
<td>22-23</td>
</tr>
<tr>
<td>SIGLPG</td>
<td>3</td>
<td>627</td>
<td>0</td>
<td>627</td>
<td>21.57</td>
<td>5.83</td>
<td>17-18</td>
</tr>
<tr>
<td>SIGLPH</td>
<td>7</td>
<td>636</td>
<td>0</td>
<td>636</td>
<td>22.03</td>
<td>6.25</td>
<td>19-20</td>
</tr>
<tr>
<td>SIGLPJ</td>
<td>9</td>
<td>919</td>
<td>1</td>
<td>660</td>
<td>24.09</td>
<td>8.64</td>
<td>20-21</td>
</tr>
<tr>
<td>SIGLPK</td>
<td>9</td>
<td>1,323</td>
<td>1</td>
<td>675</td>
<td>24.06</td>
<td>5.21</td>
<td>22-23</td>
</tr>
<tr>
<td>SIGLPL</td>
<td>9</td>
<td>2,174</td>
<td>2</td>
<td>654</td>
<td>24.30</td>
<td>7.76</td>
<td>21-22</td>
</tr>
<tr>
<td>SIGLPM</td>
<td>9</td>
<td>1,445</td>
<td>1</td>
<td>675</td>
<td>24.83</td>
<td>7.75</td>
<td>21-22</td>
</tr>
<tr>
<td>SIGLPN</td>
<td>9</td>
<td>1,189</td>
<td>2</td>
<td>846</td>
<td>30.85</td>
<td>8.62</td>
<td>21-22</td>
</tr>
<tr>
<td>SIGLPO</td>
<td>11</td>
<td>621</td>
<td>0</td>
<td>621</td>
<td>21.91</td>
<td>5.51</td>
<td>26-27</td>
</tr>
</tbody>
</table>

* Coding DNA Sequence
Smallest/lowest and largest/highest values are indicated in bold
Results and Discussion

Genomic organization and sequence analysis of the GLP gene family members in tomato

In the present study, 15 tomato GLPs with predicted oxalate oxidase activity were identified using the tomato genome sequence available in the SolGenomics database. Analysis of the corresponding genes was carried out to find out the genomic organization of these 15 genes in tomato genome. It was found that these 15 genes are located on 5 different chromosomes in tomato. Among the 5 chromosomes, chromosome 1 contains 6 GLP genes (annotated in this study as SIGLPA, SIGLPB, SIGLPC, SIGLPD, SIGLPE and SIGLPF) in a ~8.5 Mb spanning region. All the genes, except SIGLPF were found to be located in reverse orientation. Chromosome 3 was found to contain 1 GLP gene (SIGLPG) in reverse orientation and chromosome 7 was found to contain 2 GLP genes (SIGLPH and SIGLPI) in forward and reverse orientation, respectively. On the other hand, 5 GLP genes (SIGLPJ, SIGLPK, SIGLPL, SIGLPM and SIGLPN) were observed to be very closely located on chromosome 9, all in reverse orientation, in a ~17.2 kb spanning region. The SIGLPO gene was found to be present in forward orientation on chromosome 11. The schematic image of the genomic organization of all these 15 genes is presented in Figure 1A.

Among the 15 genes, SIGLPK was found to be the longest (2,174 bp), whereas SIGLPB was found to be smallest (546 bp) (Table 1). Except SIGLPG, SIGLPH and SIGLPO, all the genes were found to contain introns. The SIGLPK and SIGLPM genes contained 2 introns each, whereas the rest 10 genes were observed to possess single intron (Figure 1A). Among the introns, the 2nd intron present in the SIGLPK gene was found to be the largest (1,269 nt, data not shown). The Coding DNA Sequence (CDS) was longest in case of the SIGLPM gene (846 bp) and smallest (405 bp) in case of the SIGLPB gene (Table 2).
In silico analysis of the germinlike protein

Amino acid sequence-based analysis of tomato GLPs

The predicted molecular weight of the 15 GLPs of tomato ranged from 14.86 kDa (SIGLPB) to 30.85 kDa (SIGLPM) (Table 1). The predicted theoretical pI of the proteins ranged from 5.21 (SIGLPJ) to 9.21 (SIGLPA). Except SIGLPA and SIGLPB, signal peptides were predicted for all the proteins, where longest (26 amino acids) signal peptide was predicted in case of SIGLPO and smallest (17 amino acids) signal peptide was predicted for SIGLPG.

Amino acid sequences of these 15 GLPs were subjected to multiple sequence alignment. It was found that, in spite of having N-terminal and C-terminal unique sequences, all these proteins contained the conserved amino acid residues designated as the Box B and Box C, characteristic to plant GLPs (Figure 1B). However, the Box A was found to have a less extent of sequence
conservation in all the tomato GLPs. Apart from SlGLPK, variation in the length of the inter-motif hinge region between the two boxes (Box B and Box C) in all the tomato GLPs were found to be not more than 6 amino acids. However, the hinge region of SlGLPK was found to be significantly smaller and almost absent (Figure 1B).

In order to classify the tomato GLPs, sequence comparison was done with GLPs from barley. Six GLPs from Hordeum vulgare L. (HvGER1a, HvGER2a, HvGER3a, HvGER4c, HvGER5a and HvGER6a), belonging to separate classes (GER1, GER2, GER3, GER4, GER5 and GER6) were incorporated during cluster analysis. Among the 15 tomato GLPs, SlGLPG and SlGLPH were found to share maximum amount of sequence similarity with HvGER2a and hence they were classified as GER2 sub-family members (Figure 1C). On the other hand, the SIGLPD, SIGLPE and SIGLPF proteins shared significant sequence similarity with HvGER4c and were classified in the GER4 sub-family. Among rest of the tomato GLPs, SIGLPA, SIGLPB, SIGLPC, SIGLPJ, SIGLPK, SIGLPL, SIGLPM and SIGLPN belonged to a cluster that was found to have sequence similarity with HvGER1a. Hence, they were classified in the GER1 sub-family. The SIGLPJ and SIGLPO proteins failed to show significant sequence similarity with the reported Hordeum GLPs and hence remained unclassified in this study.

Figure 2: Comparative view of the structures of the 15 tomato GLPs with predicted oxalate oxidase activity. Structural distinctness along with the basic conserved ‘jelly roll’ structure, characteristic feature of plant GLPs, should be noted.
Structural modeling and analysis of tomato GLPs

Homology-based structural models of all these 15 tomato GLPs were prepared and analyzed in this study. Except SlGLPL, all the other GLP models were prepared using the solved crystal structure of Hordeum vulgare oxalate oxidase protein (PDB ID: 1fi2A). For the SlGLPL model, solved crystal structure of Amaranthus 11s proglobulin seed storage protein (PDB ID: 3qacA), which shared 21.43% sequence identity with SlGLPL (Table 2) was used as template. All the other GLPs shared 21.48% (SlGLPO) to 48.30% (SlGLPD and SlGLPE) sequence identity with the Hordeum vulgare oxalate oxidase protein. Length of the stretch of amino acid residues modeled was found to be highest in case of SlGLPN (207 amino acids) and lowest in case of SlGLPB (131 amino acids) (Table 2). All the tomato GLPs were found to have distinguishable structural models, having similarity with the characteristic β ‘jelly roll’ structure of plant GLPs (Figure 2). The root mean square deviations of the α-carbon backbone (R.M.S.D Cα, indicator of the structural similarity between the model and its template) of the tomato GLP structural models with their corresponding template were analyzed through structural superimposition. The R.M.S.D Cα value was found to be lowest (0.072) in case of SlGLPE and highest (4.594) in case of SlGLPL (Table 2). Interestingly, all the tomato GLP homology models, except SlGLPA, SlGLPB and SlGLPL, were observed to contain Mn2+ as ligand (Table 2).

Fig.3: A. Structural super-imposition of the 15 tomato GLPs to reveal the conserved β barrel. B. Conserved His and Glu residue(s) in the Box B and Box C of tomato GLPs, documented to be involved in Mn2+ binding. Change in these amino acids, as observed in case of SlGLPA and SlGLPB are shown inside small circles. C. View of the active site architecture, constituted by the conserved His and Glu residues’ side-chains, in Mn2+ binding tomato GLPs. Position of Mn2+ is indicated as sphere.
Analysis of the Mn$^{2+}$ binding active sites of tomato GLPs

Structural superimposition of all the tomato GLP models revealed the presence of the conserved $\beta$ barrel in these proteins (Figure 3A). From the comparison of amino acid residues present in Box B and Box C, it was found that the predicted amino acid residues (2 His and 1Glu residue in Box B; 1 His residue in Box C) involved in metal ion binding to be highly conserved among the tomato GLPs (Figure 3B). However, in case of SIGLPA, the first conserved His residue of Box B was found to be replaced by Tyr (Figure 3B, upper part) and in case of SIGLPB, the conserved His residue of Box C was observed to be replaced by Asn (Figure 3B, lower part). Orientation of the side-chains of the 3 His and 1Glu residues in the conserved $\beta$ barrel portion of the Mn$^{2+}$ binding SIGLPC protein was found to form the active site pocket suitable for Mn$^{2+}$ binding (Figure 3C). But, the amino acid replacements in case of SIGLPA and SIGLPB were found to alter the architecture of the active site for Mn$^{2+}$ binding. Comparative study of the active site architecture, in reference to the Mn$^{2+}$ binding SIGLPC protein revealed that in case of SIGLPA, projection of the Tyr side-chain is supposed to reduce the space required for binding Mn$^{2+}$ (Figure 4A). On the other hand, projection of the Asn side-chain in the active site of SIGLPB is supposed to reduce the required compactness of the active site for Mn$^{2+}$ binding (Figure 4A).

Figure 4: A. Comparative view of the active site architecture in Mn$^{2+}$ binding (SIGLPC) and not binding (SIGLPA, SIGLPB and SIGLPL) tomato GLPs. Change in architecture due to the changed or differentially projected amino acid side chain is indicated by arrow marks. B. Identification of the corresponding point mutation (C to T) in SIGLPA (right), leading to the change in amino acid residue (H to Y, left) in the active site, shown inside boxes. C. Identification of the corresponding point mutation (C to A) in SIGLPB (right), leading to the change in amino acid residue (H to N, left) in the active site, shown inside boxes.
Interestingly, SlGLPL was found to contain all the conserved amino acid residues required for Mn\(^{2+}\) binding; but the side chain projection of the conserved second His residue of Box B of SlGLPL was found to be highly altered in comparison to that of the Mn\(^{2+}\) binding SlGLPC protein (Figure 4A). However, keeping the limitations of homology-based structural modeling in mind, Mn\(^{2+}\) binding potential of SlGLPL might be examined in wet-lab condition in order to make confirm comment in this regard.

Comparison of the CDS region of SlGLPA and SLGLPB with that of SlGLPC revealed the point mutations resulting in the replacement of the aforementioned conserved amino acid residues. In case of SlGLPA, one C to T point mutation (transition) resulted in His to Tyr replacement in the Box B (Figure 4B). On the other hand, one C to A point mutation (transversion) caused the His to Asn replacement in the Box C of SlGLPB (Figure 4C). These point mutations might further be explored to design specific primers for checking the expression profile of Mn\(^{2+}\) binding and non-binding GLPs of tomato under different physiological conditions.

**Conclusion**

In the present study, the *in silico* analysis of 15 tomato GLPs with predicted oxalate oxidase activity has been documented. Through homology-based structural modeling and analyses of the tomato GLPs, their structural conservation has been depicted. Out of the 15 proteins, 3 were predicted not to bind metal ion (Mn\(^{2+}\)) and might have some other distinct physiological role. Through analysis of the active site architecture, attempt has been made to explain the differential binding of Mn\(^{2+}\) by these proteins. At the same time, the point mutations leading to the alteration in the metal binding active site of 2 (SIGLPA and SIGLPB) of the 3 proteins predicted not to bind Mn\(^{2+}\) has been identified. Hence, this work might serve as the platform for further molecular biological works targeting the functional characterization of the tomato GLPs.

**Acknowledgements**

The author thanks Science and Engineering Research Board, Department of Science and Technology, Government of India for financial support in terms of Young Scientist Award (Code: SB/YS/LS-74/2013). The author also thanks Prof. R.N. Sharma for providing experimental supports in the Department of Plant Breeding and Genetics, Bihar Agricultural College, Bihar Agricultural University, Sabour, Bhagalpur, Bihar. Helps from Dr. Mankesh Kumar, Dr. Satyendra and Dr. S.P. Singh are also acknowledged.

**References**


Christensen, A.B., H. Thordal-Christensen, G. Zimmermann, T. Gjetting, M.F. Lyngkjaer, R. Dudler, and P. Schweizer. 2004. The germin-like protein GLP4 exhibits superoxide dismutase activity and is an important component of...


