Identification of the conserved domains of ADP-Glucose Pyrophosphorylase (AGPase) protein in sweetpotato (*Ipomoea batatas* (L.) Lam.) and its two wild relatives

Hualin Nie  
University of Seoul  
Sujung Kim  
University of Seoul  
Hohyun Kim  
University of Seoul  
Ji-Seong Kim  
University of Seoul  
Sun-Hyung Kim  
University of Seoul

Keywords: ADP-glucose pyrophosphorylase, conserved domain, subunit, sweetpotato

Posted Date: March 8th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-277143/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

The conserved domains are defined as recurring units in molecular evolution, which are commonly used to interpret the molecular function and biochemical structure of proteins. The AGPase amino acid sequences of three species from the *Ipomoea* genus were identified to investigate their physicochemical and biochemical characteristics. The molecular weights (MW), isoelectric point (pI), instability index (II), and grand average of hydropathy (GRAVY) showed considerable differences in each plant. The aliphatic index (AI) values of sweetpotato AGPase proteins were higher in the small subunit than in the large subunit. The AGPase proteins from sweetpotato contain an LbH_G1P_AT_C domain in the C-terminal region and various domains (NTP_transferase, ADP_Glucose_PP, or Glyco_tranf_GTA) in the N-terminal region. On the other hand, most of its two relatives (*I. trifida* and *I. triloba*) only contain the NTP_transferase domain in the N-terminal region. These findings suggested that these conserved domains were species specificity and related to the subunit types of AGPase proteins. The study may enable research on the AGPase-related specific characteristics of sweetpotatoes, which do not exist in the other two species, such as starch metabolism and tuberization mechanism.

Introduction

ADP-glucose pyrophosphorylase (AGPase; EC: 2.7.7.27) is a regulatory enzyme that catalyzes the biosynthesis of alpha 1,4-glucans (glycogen or starch) in photosynthetic bacteria and plants (Smith-White and Preiss 1992). In higher plants, it is a heterotetramer composed of two different but closely related subunits (α2β2): “small” (α subunit, 50–54 kDa) and “large” subunits (β subunit, 51–60 kDa) based on the size difference (Smith-White and Preiss 1992; Ballicora et al. 2004). The small subunit is responsible for the catalytic activity, whereas the large subunit plays regulatory roles (Crevillén et al. 2003; Ballicora et al. 2004). These subunits are necessary for the optimal activity of the native enzyme in plants; a lack of one of the subunits will reduce the activity of the AGPase and influence the synthesis of starch (Li and Preiss 1992). In sweetpotato, AGPase is a key enzyme controlling starch synthesis and is considered an important determinant of the sink activity of the roots (Yatomi et al. 1996; Tsubone et al. 2000). Many AGPase genes have been cloned and studied in sweetpotatoes (Lee et al. 2000; Seo et al. 2015; Zhou et al. 2016).

The protein domains can be considered distinct functions and structural units of proteins that are usually identified as repeating (sequence or structural) units (Ingolfsson and Yona 2008; Li et al. 2012). In molecular evolution, these domains may have been reorganized in different arrangements in protein function annotation (Ingolfsson and Yona 2008), protein structure determination (Marchler-Bauer et al. 2012), and protein engineering (Guerois and Serrano 2001). Conserved domains are defined by a conserved domain database (CDD) as repeating units in molecular evolution, the extent of which can be determined by sequence and structural analysis (Marchler-Bauer et al. 2012).

Sweetpotato (*Ipomoea batatas* (L.) Lam.) is a hexaploid (2x = 6n = 90) perennial tuberization crop belonging to the family Convolvulaceae (Welbaum 2015). Two non-tuberization diploid *Ipomoea* species, *I. trifida* (H.B.K.) G. Don (2n = 2x = 30) and *I. triloba* L. (2n = 2x = 30), have been reported to be the putative progenitors of sweetpotato, which are commonly considered to be model species for sweetpotato research (Roullier et al. 2013; Wu et al. 2018). In this study, the AGPase genes were screened from sweetpotato and its two related species to investigate the conserved domains of the coding protein. The differences in these domains can be used to confirm the functions of the AGPase protein between the sweetpotato and its two relatives.

Methods

Identification of AGPase amino acid sequences
Sweetpotato Genomics Resource (http://sweetpotato.plantbiology.msu.edu/index.shtml) and NCBI databases (https://www.ncbi.nlm.nih.gov/) were used to identify the AGPase domain-containing proteins in the three species. The amino acid sequence of the AGPase protein *IbAGPa1* (BAF47744.2) was used as the driver sequence for BLAST-search.

The ProtParam (http://www.expasy.org/tools/protparam.html) of ExPASy (Expert protein analysis system, https://www.expasy.org/) tool was used to compute the physicochemical characteristics of AGPase proteins in the three species, including the number of amino acids, molecular weight, theoretical isoelectric point (pI), instability (II) and aliphatic index (AI), and grand average of hydropathy (GRAVY) (Gasteiger et al. 2005).

**Multiple-sequence alignment and phylogenetic tree structure**

The amino acid sequences of the AGPase proteins in FASTA formats were used for multiple-sequence alignment using the CLC Sequence Viewer 7.6 software (CLC bio, Aarhus, Denmark). A neighbor-joining phylogenetic tree was constructed using MEGA X 10.1 software (Pennsylvania State University, US) with the following parameters: bootstrap analysis of 1,000 replicates, Poisson correction method, and pairwise deletion (Kumar et al. 2018).

**Conserved domain analysis**

Pfam (http://pfam.janelia.org/), SMART (http://smart.embl-heidelberg.de/), and CDD (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) were used to explore the conserved domains of the AGPase proteins. The selected conserved domains were drawn using DOG 2.0.1 software (Ren et al. 2009).

**Results**

**Identification of AGPase proteins**

Forty-five AGPase domain-containing proteins from *I. batatas* (26 accessions), *I. tridra* (10 accessions), and *I. triloba* (9 accessions) were identified and used for various analyses (Table 1). The sizes of these proteins were distinctly different; the amino acids ranged from 165 to 525 and the molecular weights (MW) ranged from 18.35 to 58.19 kDa.

The isoelectric point (pl), which represents the average pH of the molecule without a net electrical charge or electrically neutrality, was 4.71–9.53 in all categories. The average pl of *I. batatas*, *I. trifida*, and *I. triloba* AGPase were 6.83, 7.11, and 6.47, respectively. The instability index (II), which represents the stability and instability of a polypeptide at ≤ 40 and > 40, respectively, indicated 40 or less in AGPase of *I. batatas*. In contrast, some AGPases of the *I. trifida* and *I. triloba* were 40 or more. The aliphatic index (AI), which represents the relative volume of the aliphatic side chains of a polypeptide, was similar in the three species, but there were differences between subunits of *I. batatas* AGPase. Higher AI values were observed for the small subunits than the large subunits of *I. batatas* AGPase. The grand average of hydropathy (GRAVY), which was analyzed to determine the hydropathy of AGPase, showed that *I. batatas* had different characteristics from the other two species. All *I. batatas* AGPases showed negative values, whereas some of the *I. trifida* and *I. triloba* AGPases had positive values.

**Conserved domain analysis**

Six types of conserved domains that showed different distributions were included in the AGPase proteins of these three species (Fig 1b, Supplementary Table 1). Most of the *I. trifida* and *I. triloba* AGPase had only the NTP_transferase domain and some had two conserved domains: NTP_transferase at the N-terminal and Hexapep or Cpn60_TCP1 at the C-terminal. On the other hand, the *I. batatas* AGPase proteins had four types of conserved domains (NTP_transferase, LbH_G1P_AT_C, ADP_Glucose_PP, and Glyco_tranf_GTA_type); each of them had two conserved domains. All of the *I. batatas* AGPase proteins had the LbH_G1P_AT_C domain at the C-terminals, but the N-terminals differed according to
the subunit. The N terminal of all large subunits of *I. batatas* AGPase proteins has the NTP_transferase domain only except for CAB51610.1, whereas all small subunits have ADP_Glucose_PP domain except for CAB55496.1, AAA19648.1, and CAA86726.1. The proteins with this exception all had partial sequences and had the Glyco_tranf_GTA_type domain at the C-terminals instead.

### Phylogenetic analysis

The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei 1987). Fig 1a presents the optimal tree with the sum of the branch length = 29.09. This analysis involved 45 amino acid sequences and 512 positions. The conserved domains were labeled on the amino acid sequences (Fig 1a). The length and type of the domain were different for each species. Based on the phylogenetic tree, AGPase proteins from these species were classified into two large subunit groups and two small subunit groups.

### Discussion

AGPase is an important factor involved in the tuberous root of sweetpotatoes because it is a vital enzyme in starch synthesis (Yatomi et al. 1996; Tsubone et al. 2000). Although it is also present in *I. trifida* and *I. triloba*, as well as in plants of the genus *Ipomoea*, they all have different physiological properties from sweetpotatoes, such as non-tuberization. Therefore, AGPase is believed to have different structures or different functions in plants of the genus *Ipomoea*. The AGPase identification of sweetpotatoes and two non-tuberous *Ipomoea* species performed in this study is very important for understanding the relationship between plants of the genus *Ipomoea* and the functions of each species.

Sweetpotato is a polyploid crop of *I. trifida*, but it is unclear if it is autopolyploid or allopolyploid (Roullier et al. 2013; Wu et al. 2018). The amount of AGPases increased by whole-genome duplication in sweetpotatoes from its relatives. This result is consistent with a study showing that the number of *rboh* genes in the polyploid plant, *Gossypium hirsutum*, was higher than its progenitor plants *G. raimondii* and *G. arboreum* (Wang et al. 2020). Moreover, some AGPases in *I. trifida* and *I. triloba* exhibited an II value ≥ 40, which means an unstable state, but there was no AGPase representing an II value ≥ 40 in *I. batatas* (Table 1). This suggests that some of the genes that were unstable during the evolution of *I. batatas* may have been deleted.

A difference in the domain composition of AGPase was observed between sweetpotatoes and the other *Ipomoea* plants; *I. batatas* has a more complex composition (Fig. 1b). The N-terminal of the small subunit and the C-terminal in sweetpotatoes were composed differently from the domains of the two species. These results suggest that LbH_G1P_AT_C at the C-terminal and ADP_Glucose_PP and Glyco_tranf_GTA_type at the N-terminal of the small subunit contribute to the different functions and regulations than non-tuberous relative plants. Many studies have shown that genes can be orthologs or paralogs by domain architectures, such as the insertion and deletion of new domains during evolution (Björklund et al. 2005; Forslund et al. 2011). Although this study cannot confirm the homolog genes of each AGPase in the genus *Ipomoea* plants, the evolutionary process of the genome among these plants, including AGPase, is expected to be revealed through further studies.

### Conclusion

Sweetpotato AGPases has relatively conserved domains compared to *I. trifida* and *I. triloba*. The small subunit of AGPase showed complex structures in sweetpotatoes compared to the other two species. Sweetpotato AGPase had the LbhH_G1P_AT_C domain in the C-terminal region, which was not present in *I. trifida* and *I. triloba*. This suggests that the structure of AGPase in sweetpotato, which is different from the other two species, plays important roles in certain
functions of sweetpotatoes, such as starch biosynthesis and tuber formation. More isolation studies and further examination of gene expression will be needed to clarify the functional role of sweetpotato-specific domains in tuberization.

Declarations

Funding

Not applicable

Conflict of interest

We declare that we have no conflict of interest.

Availability of data and material

Not applicable

Code availability

Not applicable

Author contributions

All authors contributed to the study conception and design. Kim SH conceived the original research plan; Nie H performed the data collection and wrote the manuscript; Kim SJ, Kim HH, and Kim JS revised the manuscript. All authors have reviewed and approved the final version of the manuscript.

References


Table 1. Biochemical and physicochemical characteristics of AGPase proteins in the three species.
<table>
<thead>
<tr>
<th>Species</th>
<th>Accession No.</th>
<th>Subunit</th>
<th>Amino acids</th>
<th>Molecular weight (WM)</th>
<th>Isoelectric point (pI)</th>
<th>Instability index (II)</th>
<th>Aliphatic index (AI)</th>
<th>Grand average of hydropathy (GRAVY)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. batatas</td>
<td>BAF47744.2</td>
<td>Small</td>
<td>522</td>
<td>57155.24</td>
<td>6.74</td>
<td>39.79</td>
<td>91.24</td>
<td>-0.178</td>
</tr>
<tr>
<td>I. batatas</td>
<td>AFL55400.1</td>
<td>Small</td>
<td>522</td>
<td>57143.19</td>
<td>6.74</td>
<td>39.50</td>
<td>90.48</td>
<td>-0.188</td>
</tr>
<tr>
<td>I. batatas</td>
<td>AAS66988.1</td>
<td>Small</td>
<td>522</td>
<td>57188.32</td>
<td>6.74</td>
<td>39.42</td>
<td>91.23</td>
<td>-0.166</td>
</tr>
<tr>
<td>I. batatas</td>
<td>AAA19648.1</td>
<td>Small</td>
<td>303</td>
<td>33530.51</td>
<td>5.52</td>
<td>35.06</td>
<td>96.30</td>
<td>-0.129</td>
</tr>
<tr>
<td>I. batatas</td>
<td>CAA86726.1</td>
<td>Small</td>
<td>302</td>
<td>33374.32</td>
<td>5.39</td>
<td>35.14</td>
<td>96.62</td>
<td>-0.115</td>
</tr>
<tr>
<td>I. batatas</td>
<td>CAA58473.1</td>
<td>Small</td>
<td>427</td>
<td>47300.22</td>
<td>6.13</td>
<td>36.29</td>
<td>97.12</td>
<td>-0.119</td>
</tr>
<tr>
<td>I. batatas</td>
<td>AFL55401.1</td>
<td>Small</td>
<td>523</td>
<td>57164.19</td>
<td>8.02</td>
<td>37.38</td>
<td>90.15</td>
<td>-0.194</td>
</tr>
<tr>
<td>I. batatas</td>
<td>BAF47745.1</td>
<td>Small</td>
<td>523</td>
<td>57178.21</td>
<td>8.02</td>
<td>37.38</td>
<td>90.34</td>
<td>-0.190</td>
</tr>
<tr>
<td>I. batatas</td>
<td>AAS66987.1</td>
<td>Small</td>
<td>523</td>
<td>57179.24</td>
<td>8.02</td>
<td>36.64</td>
<td>90.52</td>
<td>-0.183</td>
</tr>
<tr>
<td>I. batatas</td>
<td>AFL55399.1</td>
<td>Large</td>
<td>525</td>
<td>58055.43</td>
<td>8.92</td>
<td>34.29</td>
<td>88.44</td>
<td>-0.164</td>
</tr>
<tr>
<td>I. batatas</td>
<td>AGB85112.1</td>
<td>Large</td>
<td>525</td>
<td>57990.31</td>
<td>8.82</td>
<td>33.14</td>
<td>87.80</td>
<td>-0.158</td>
</tr>
<tr>
<td>I. batatas</td>
<td>BAF47749.1</td>
<td>Large</td>
<td>525</td>
<td>58117.46</td>
<td>8.93</td>
<td>35.26</td>
<td>87.50</td>
<td>-0.164</td>
</tr>
<tr>
<td>I. batatas</td>
<td>AFL55398.1</td>
<td>Large</td>
<td>518</td>
<td>57269.40</td>
<td>6.37</td>
<td>29.97</td>
<td>85.08</td>
<td>-0.178</td>
</tr>
<tr>
<td>I. batatas</td>
<td>BAF47748.1</td>
<td>Large</td>
<td>518</td>
<td>57269.36</td>
<td>6.25</td>
<td>29.73</td>
<td>85.08</td>
<td>-0.177</td>
</tr>
<tr>
<td>I. batatas</td>
<td>AGB85111.1</td>
<td>Large</td>
<td>517</td>
<td>57376.52</td>
<td>6.41</td>
<td>28.99</td>
<td>84.29</td>
<td>-0.190</td>
</tr>
<tr>
<td>I. batatas</td>
<td>AFL55396.1</td>
<td>Unknown</td>
<td>517</td>
<td>57577.74</td>
<td>7.01</td>
<td>35.32</td>
<td>86.36</td>
<td>-0.245</td>
</tr>
<tr>
<td>I. batatas</td>
<td>BAF47746.1</td>
<td>Large</td>
<td>517</td>
<td>57616.78</td>
<td>6.69</td>
<td>36.61</td>
<td>87.31</td>
<td>-0.234</td>
</tr>
<tr>
<td>I. batatas</td>
<td>CAB52196.1</td>
<td>Unknown</td>
<td>450</td>
<td>50090.21</td>
<td>5.38</td>
<td>35.94</td>
<td>89.04</td>
<td>-0.168</td>
</tr>
<tr>
<td>I. batatas</td>
<td>BAF47747.1</td>
<td>Large</td>
<td>515</td>
<td>57562.13</td>
<td>7.08</td>
<td>31.74</td>
<td>88.99</td>
<td>-0.204</td>
</tr>
<tr>
<td>I. batatas</td>
<td>AFL55397.1</td>
<td>Large</td>
<td>515</td>
<td>57485.94</td>
<td>6.44</td>
<td>32.78</td>
<td>88.80</td>
<td>-0.194</td>
</tr>
<tr>
<td>I. batatas</td>
<td>AGB85109.1</td>
<td>Large</td>
<td>517</td>
<td>57527.64</td>
<td>6.44</td>
<td>37.97</td>
<td>87.50</td>
<td>-0.237</td>
</tr>
<tr>
<td>Species</td>
<td>Accession no.</td>
<td>Type</td>
<td>Length</td>
<td>Coverage</td>
<td>NTM</td>
<td>Identity</td>
<td>Support</td>
<td>Score</td>
</tr>
<tr>
<td>-------------</td>
<td>---------------</td>
<td>----------</td>
<td>--------</td>
<td>----------</td>
<td>-------</td>
<td>----------</td>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td><em>I. batatas</em></td>
<td>CAB55495.1</td>
<td>Unknown</td>
<td>490</td>
<td>54707.53</td>
<td>7.14</td>
<td>36.97</td>
<td>89.33</td>
<td>-0.227</td>
</tr>
<tr>
<td><em>I. batatas</em></td>
<td>AGB85110.1</td>
<td>Large</td>
<td>515</td>
<td>57559.03</td>
<td>6.31</td>
<td>31.13</td>
<td>89.55</td>
<td>-0.212</td>
</tr>
<tr>
<td><em>I. batatas</em></td>
<td>AAC21562.1</td>
<td>Large</td>
<td>517</td>
<td>57686.94</td>
<td>7.55</td>
<td>38.55</td>
<td>86.92</td>
<td>-0.234</td>
</tr>
<tr>
<td><em>I. batatas</em></td>
<td>CAB55496.1</td>
<td>Large</td>
<td>385</td>
<td>43443.49</td>
<td>5.35</td>
<td>32.30</td>
<td>85.82</td>
<td>-0.224</td>
</tr>
<tr>
<td><em>I. batatas</em></td>
<td>CAB51610.1</td>
<td>Large</td>
<td>306</td>
<td>34636.48</td>
<td>5.13</td>
<td>37.96</td>
<td>86.63</td>
<td>-0.300</td>
</tr>
</tbody>
</table>

**Figures**
Figure 1

Phylogenetic tree and domain structure of the AGPase proteins in I. batatas (black circles), I. tritida (red quadrangles), and I. triloba (green triangles). The numbers at the nodes indicate the bootstrap values. The conserved domains are indicated by colored blocks on the right. Gray, NTP transferase; green, LbH_G1P_AT_C; blue, Glyco_tranf_GTA_type; purple, Hexapep; red, Cpn60_TCP1; orange, ADP_Glucose_PP

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1.xlsx