BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 9(11)

Comparative computational analysis of ADP Glucose Pyrophosphorylase in plants

Saroj Rani[1, 2], Pradeep Sharma[1]*, Anil Sharma[2] & Ravish Chatrath[3]

[1]Biotechnology Unit, Directorate of Wheat Research Karnal, Haryana, India; [2]Department of Biotechnology, Maharishi Markandeshwar University Mullana, Ambala, India; [3]Crop Improvement Division, Directorate of Wheat Research Karnal, Haryana, India; Pradeep Sharma – Email: neprads@gmail.com; *Corresponding author

Received April 04, 2013; Revised April 11, 2013; Accepted April 22, 2013; Published June 29, 2013

Abstract:
ADP-glucose pyrophosphorylase (AGPase), a key enzyme involved in higher plant starch biosynthesis, is composed of pairs of large (LS) and small subunits (SS). Ample evidence has shown that the AGPase catalyzes the rate limiting step in starch biosynthesis in higher plants. In this study, we compiled detailed comparative information about ADP glucose pyrophosphorylase in selected plants by analyzing their structural features e.g. amino acid content, physico-chemical properties, secondary structural features and phylogenetic classification. Functional analysis of these proteins includes identification of important 10 to 20 amino acids long motifs arise because specific residues and regions proved to be important for the biological function of a group of proteins, which are conserved in both structure and sequence during evolution. Phylogenetic analysis depicts two main clusters. Cluster I encompasses large subunits (LS) while cluster II contains small subunits (SS).

Key words: Computational tools, Aliphatic index, instability index, isoelectric point.

Background:
Starch is an important carbohydrate and the primary energy source in plants. It has numerous industrial applications as reviewed in Slattery *et al.* **[1, 6].** Starch biosynthesis occurs mainly by the participation of three enzymes: ADP-glucose pyrophosphorylase (AGPase), starch synthase, and branching enzymes **[2, 3].** The first enzyme in starch biosynthesis is the AGPase that catalyzes the conversion of Glc-1-P and ATP to ADP-glucose and pyrophosphate (PPi). ADP-glucose is then used by starch synthase for the synthesis of polyglucans. Many researchers have revealed that the AGPase catalyzes the rate limiting step in starch biosynthesis in higher plants **[1, 2, 4].** AGPase from higher plants has a heterotetrameric structure ($\alpha_2\beta_2$) composed of pairs of small (SS) and large (LS) subunits encoded by at least two different genes **[5].** The large subunit (LS) plays a major role in allosteric regulation through its interaction with the small catalytic subunit (SS). The LS is encoded by the shrunken-2 (Sh2) and the SS by brittle-2 (Bt2) **[7].** Both Sh2 and Bt2 genes show considerable amino acid identity (43.2%) and similarity (61%). Maize (*Zea mays*) and rice (*Oryza sativa*), the two major cereals, show 93% identity in the amino acids sequence for the LS of AGPase enzyme. In case of wheat both large and small subunit show only 49 % identity in their amino acid sequences.

Wheat, Rice, maize, barley, and potato are important amongst staple crops as these are primarily consumed by humans otherwise also these are the cheapest source of carbohydrates and proteins used as food by one third population of the world. Economically, wheat is one of the major food crops both in terms of area and production. Wheat is grown in diverse environments, from cool rain-fed to hot dry-land areas around the world. This wide spread cultivation of the crop all along the globe is largely due to high versatility of its genome. The seed number and seed weight are important yield components of wheat for determining wheat production. Particularly starch, which accounts for 65-75% of wheat grain dry weight and composed of glucan chains, amylopectin and amylase, a major

determinant of yield. This has rendered the use for comparative analysis of ADP-glucose pyrophosphorylase in selected plants. In addition, phylogenetic analysis was also performed to know the evolutionary relatedness among AGPase.

Methodology:

Protein sequence retrieval

Protein sequences of small subunit (SS) and large subunit (LS) of AGPase from wheat, rice, maize, potato and Arabidopsis were retrieved from protein database of NCBI (National Center for Biotechnology Information, (http://www.ncbi.nlm.nih.gov/protein/) in FASTA format.

Physico-chemical characterization

The ProtParam tool (http://web.expasy.org/protparam/) of ExPASy was used to compute amino acid composition (%), molecular weight, theoretical isoelectric point (pI), number of positively and negatively charged residues, extinction coefficient, instability and aliphatic index, Grand Average of Hydropathy (GRAVY).

Secondary structural properties

Secondary structural properties of the protein including alpha helix, 3_{10} helix, Pi helix, beta bridge, extended strand, beta turns, bend region, random coil, ambiguous and other states were computed by using SOPMA (Self Optimized Prediction Method with Alignment, http://npsapbil.ibcp.fr/cgibinnpsa_automat.pl?page=/NPSA/npsa_sopma.html) tool of NPS (Network Protein Sequence Analysis).

Prediction of functional properties

The motif prediction analysis was carried out with the help of Expasy's prosite tool. For functional analysis, the motifs of the AGPase protein sequences were identified by using Prosite (http://prosite.expasy.org/). Input data type was in FASTA format and motifs were scanned against prosite patterns.

Identification of Signature Logo using Web tool

Logo of AGPase was generated using Web Logo tool (http://weblogo.berkeley.edu/). In this overall height of the stack indicates the sequence conservation at that position, while the heights of the symbols within the stack indicate the relative frequency of each amino acid at that position.

Phylogenetic analysis

Twelve sequences of both large and small subunits of wheat, rice, maize, barley, potato and Arabidopsis were aligned by ClustalW tool and output file of this program was used for generation of phylogenetic tree (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

Results & Discussion:

For all the AGPase physicochemical characterization, secondary structure properties, motif and phylogenetic analysis was carried out by using various computational tools. In maize additional tyrosine and serine residues increases the seed weight 11-18% without increasing or decreasing the percentage of starch **[8].** From ProtParam result, it was observed that for all the residues on average the percentage of serine was higher than the other residues but the percentage of tyrosine was average. The percentage of serine was higher as compare to tyrosine in all subunits analyzed however it was highest in large subunit of *Arabidopsis thaliana* and lowest in small subunit of maize. In contrast to this the percentage of tyrosine was approximately equal in all the sequences except small subunits of wheat and maize **(Figure 1**).

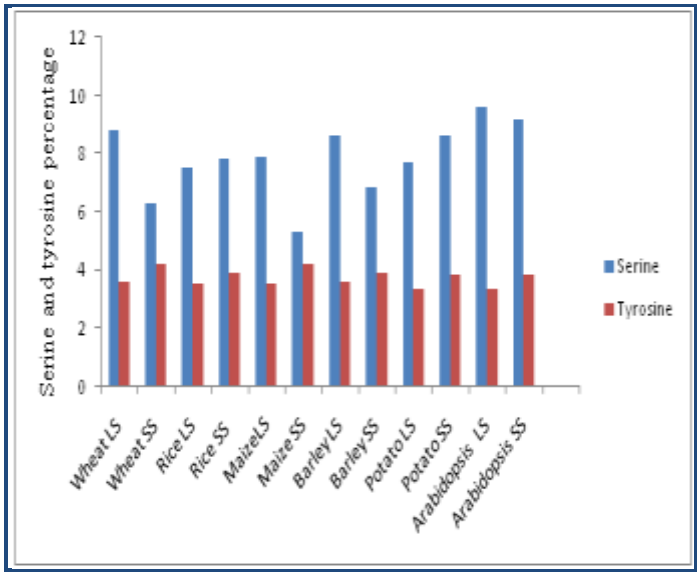

Figure 1: Serine and tyrosine percentage of AGPase in selected plants

The total number of positively (Arg + Lys) and negatively (Asp + Glu) charged residues of AGPase members were observed **Table 1 (see supplementary material)**. For all members, the total number of negatively charged residues exceeded the total number of positive charged residues except large subunit of potato and *Arabidopsis thaliana*. This possible variation might be due to their isoelectric point in acidic range. For the remaining members, the isoelectric point was within alkaline range. Extinction coefficient for all AGPase was observed higher almost with in a same range. High extinction coefficient means higher concentration of lysine, tryptophan and tyrosine. This prediction is useful to study protein-protein interaction studies. Stability of protein is described in terms of its stability index whether a protein is stable or not, can be described by its instability index. Instability index for large subunit of wheat, barley, Arabidopsis and small subunit of potato and rice is higher than 40 and thus describing these proteins unstable. It is noteworthy that high aliphatic index was observed for small subunit of all plants as compare to large subunit. The higher aliphatic index indicates higher concentration of alanine, valine, isoleucine and leucine occupying the relative volume of a protein **[9]**.

In addition to this higher aliphatic index also provides higher thermo stability. The results obtained in case of instability index and aliphatic index were contradictory while compared for AGPase analysis in rice, potato and Arabidopsis. According to instability index, these proteins are unstable but their aliphatic index is high enough to say that they are stable. These finding are in consistent with earlier research **[8].** Grand Average of Hydropathy (GRAVY) was computed for all the members. A range of GRAVY value was observed from -0.253 to -0.131`for AGPase in selected plants. SOPMA analysis was done for all AGPase members and it showed a high value for random coil in

all the members **Table 2 (see supplementary material)** while alpha helix were found approximately equal to extended strands.

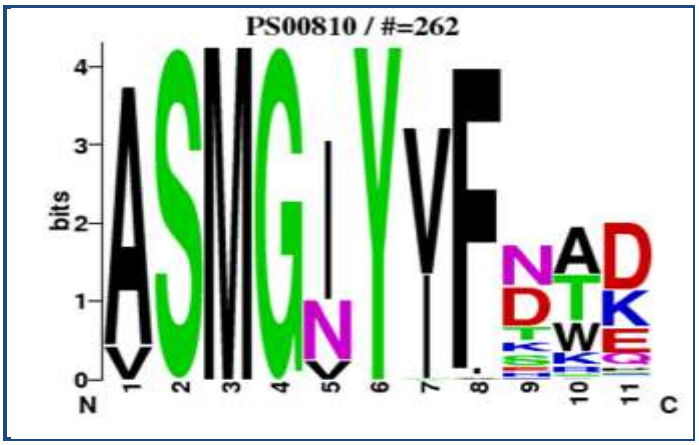

Figure 2: WebLogo representation of motif of ADP Glucose pyrophosphorylase. The amino acid type and position are shown on the *x* axis. The overall height of the amino acid stacks, plotted on the *y* axis, indicates the sequence conservation at a given position, while the height of individual symbols within a stack indicates the relative frequency of an amino acid at that position. Amino acids are color coded according to their type as basic (blue), hydrophobic (black), polar/nonpolar (green), and acidic (red).

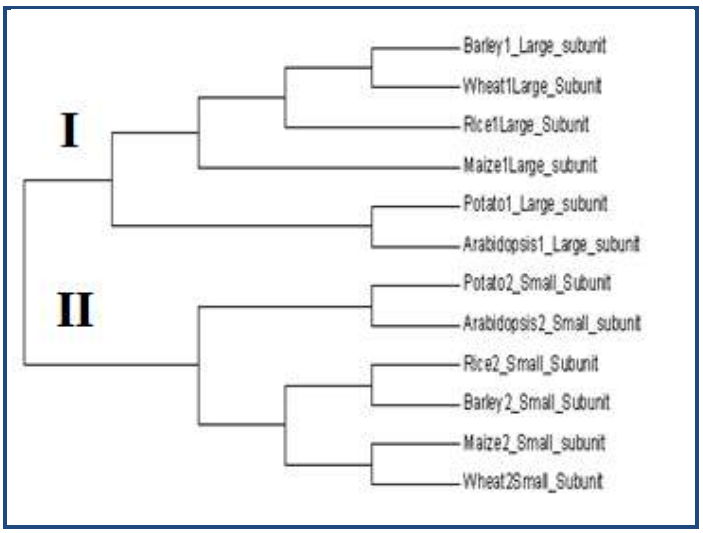

Figure 3: Phylogenetic tree of ADP Glucose Pyrophosphorylase obtained using ClustalW.

High value for random coil bears important significance in the study of protein tertiary structure and related functions. Functional analysis of these proteins includes identification of important motifs **Table 3 (see supplementary material)**. These motifs were 10 to 20 amino acids in length arise because specific residues and regions proved to be important for the biological function of a group of proteins, which are conserved in both structure and sequence during evolution. In this study, a signature logo of ADP glucose pyrophosphorylase was also generated by web logo tool. The overall height of the stack indicates the sequence conservation at that position, while the height of the symbols within the stack indicates the relative frequency of each amino acid at that position **(Figure 2)**. Phylogenetic analysis depicts two main clusters **(Figure 3).** Cluster I encompasses large subunits (LS) while cluster II contains Small subunits (SS). This study will provide a good foundation for further functional analysis of AGPase of other crops. However, the outcome of this study needs further validation by experimental approach.

Conclusion:

In this study, we compiled detailed comparative information about ADP-glucose pyrophosphorylase in selected plants by analyzing their structural features e.g. amino acid content, physico-chemical properties, secondary structural features and phylogenetic classification. Present investigation will provide an insight for the biologists working with ADP-glucose pyrophosphorylase in order to understand the functionality of AGPase.

Acknowledgment:

This work was supported by the UGC sponsored Rajiv Gandhi National Fellowship sponsored by University Grants Commission to SR, Agri Bioinformatics Promotion Program, Department of Information Technology, Ministry of Communications & Information Technology, Government of India, New Delhi to RC and Indian Council of Agricultural research of New Delhi, India to PS. The authors are thankful to Project Director, DWR, Karnal for providing facilities for this work.

References:

[1] Gasteiger E *et al.* Protein Identification and Analysis Tools on the ExPASy Server. 2005 In: John M Walker ed, The Proteomics Protocols Handbook, Humana Press 571-607
[2] Kyte J & Doolottle RF, *J Mol Biol.* 1982 **157**: 105 [PMID: 7108955]
[3] Martin C & Smith AM, *Plant Cell.* 1995 **7:** 971 [PMID: 7640529]
[4] Salamone PR *et al. FEBS Lett.* 2000 **482**: 113 [PMID: 11018533]
[5] Tuncel A *et al. Biophysical J.* 2008 **95:** 3628-3639 [PMID: 18641076]
[6] Slattery CJ *et al. Trends Plant Sci.* 2000 **5**: 291 [PMID: 10871901]
[7] Bhave MR, *et al. Plant Cell.* 1990 **2**: 581 [PMID: 1967077]
[8] Giroux MJ *et al. Proc Natl Acad Sci U S A.* 1996. **93**: 5824 [PMID: 8650177]
[9] Saha A *et al. Bioinformation.* 2013 **9**: 42 [PMID: 3563415]

Edited by P Kangueane
Citation: **Rani *et al.*** Bioinformation 9(11): 572-576 (2013)

BIOINFORMATION open access

Supplementary material:

Table 1: Various parameters computed using Expasy's ProtParam tool of AGPase in selected plants

Protein	Accession no	Length	Molecular weight	pI	-R	+R	EC	II	AI	Gravy
Wheat (LS)	P12299	522	57808.7	6.12	65	60	45435	42.86	80.52	-0.253
Wheat (SS)	P30523	473	52147.4	5.54	62	50	46550	35.91	90.76	-0.226
Rice(LS)	Q688T8	519	57653.7	6.34	63	59	44070	35.72	82.85	-0.212
Rice(SS)	P15280	514	56104.0	6.58	59	57	46675	42.56	90.37	-0.159
Maize (LS)	P55241	516	57071.0	6.16	60	55	44195	38.88	85.06	-0.207
Maize (SS)	Q947C0	475	52197.7	5.46	59	50	52050	33.35	91.87	-0.164
Barley (LS)	P30524	523	57932.6	6.17	65	60	45560	43.13	80.75	-0.242
Barley (SS)	P55238	513	56049.2	6.11	60	55	46675	36.52	91.72	-0.118
Potato (LS)	P55243	483	53602.3	8.92	55	62	46090	36.44	88.03	-0.214
Potato (SS)	P23509	521	57240.3	6.73	61	60	46675	44.18	91.21	-0.196
Arabidopsis (LS)	P55229	522	57673.8	8.02	61	63	47705	42.22	92.87	-0.174
Arabidopsis (SS)	P55228	520	56650.5	6.13	59	55	46675	34.86	93.04	-0.131

PI- isoelectric point; R- number of negatively & positively charged residue; EC- extinction coefficient; II-instability index; AI-aliphatic index, Gravy-Grand average of hydropathicity

Table 2: Secondary structure prediction of AGPase through SOPMA

Protein	α Helix	3_{10} Helix	Pi Helix	β Bridge	Extended Strand	β Turn	Bend Region	Random Coil	Ambiguous States	Other States
Wheat(LS)	24.33	0.00	0.00	0.00	22.03	7.28	0.00	46.36	0.00	0.00
Wheat(SS)	22.62	0.00	0.00	0.00	22.41	6.77	0.00	48.20	0.00	0.00
Rice(LS)	25.24	0.00	0.00	0.00	21.97	6.55	0.00	46.24	0.00	0.00
Rice (SS)	25.29	0.00	0.00	0.00	22.18	7.20	0.00	45.33	0.00	0.00
Maize (LS)	25.58	0.00	0.00	0.00	22.48	7.56	0.00	44.38	0.00	0.00
Maize (SS)	22.95	0.00	0.00	0.00	23.37	6.32	0.00	47.37	0.00	0.00
Barley (LS)	24.09	0.00	0.00	0.00	22.56	7.27	0.00	46.08	0.00	0.00
Barley (SS)	25.93	0.00	0.00	0.00	22.61	6.63	0.00	44.83	0.00	0.00
Potato(LS)	24.84	0.00	0.00	0.00	24.02	7.04	0.00	44.10	0.00	0.00
Potato(SS)	24.57	0.00	0.00	0.00	22.46	5.37	0.00	47.60	0.00	0.00
Arabidopsis(LS)	24.71	0.00	0.00	0.00	23.18	6.90	0.00	45.21	0.00	0.00
Arabidopsis (SS)	25.00	0.00	0.00	0.00	24.81	6.54	0.00	43.65	0.00	0.00

Table 3: Motif prediction of AGPase in selected plants

Protein	Motif ID	Motif description	Start	End	Pattern
Wheat	PS00808	ADP_GLC_PYROPHOSPH_1	96	115	GGGtGTqLfpLTstrAtPAV
(LS)	PS00809	ADP_GLC_PYROPHOSPH_2	185	193	WFrGTADAV
	PS00810	ADP_GLC_PYROPHOSPH_3	301	311	ASMGVYVFkrD
Wheat	PS00808	ADP_GLC_PYROPHOSPH_1	49	68	GGGaGTrLypLTkkrAkPAV
(SS)	PS00809	ADP_GLC_PYROPHOSPH_2	138	146	WFqGTADAV
	PS00810	ADP_GLC_PYROPHOSPH_3	250	260	ASMGIYVIskH
Rice	PS00808	ADP_GLC_PYROPHOSPH_1	93	112	GGGtGTqLfpLTstrAtPAV
(LS)	PS00809	ADP_GLC_PYROPHOSPH_2	182	190	WFqGTADAV
	PS00810	ADP_GLC_PYROPHOSPH_3	298	308	ASMGVYVFkrD
Rice	PS00808	ADP_GLC_PYROPHOSPH_1	90	109	GGGaGTrLypLTkkrAkPAV
(SS)	PS00809	ADP_GLC_PYROPHOSPH_2	179	187	WFqGTADAV
	PS00810	ADP_GLC_PYROPHOSPH_3	291	301	ASMGIYVIskN
Maize	PS00808	ADP_GLC_PYROPHOSPH_1	90	109	GGGtGSqLfpLTstrAtPAV
(LS)	PS00809	ADP_GLC_PYROPHOSPH_2	179	187	WFqGTADSI
	PS00810	ADP_GLC_PYROPHOSPH_3	295	305	ASMGIYVFkkD
Maize	PS00808	ADP_GLC_PYROPHOSPH_1	51	70	GGGaGTrLypLTkkrAkPAV
(SS)	PS00809	ADP_GLC_PYROPHOSPH_2	140	148	WFqGTADAV
	PS00810	ADP_GLC_PYROPHOSPH_3	252	262	ASMGIYVFskD
Barley	PS00808	ADP_GLC_PYROPHOSPH_1	97	116	GGGtGTqLfpLTstrAtPAV
(LS)	PS00809	ADP_GLC_PYROPHOSPH_2	186	194	WFrGTADAV
	PS00810	ADP_GLC_PYROPHOSPH_3	302	312	ASMGVYVFkrD

	PS00808	ADP_GLC_PYROPHOSPH_1	89	108	GGGaGTrLypLTkkrAkPAV
Barley	PS00809	ADP_GLC_PYROPHOSPH_2	178	186	WFqGTADAV
(SS)	PS00810	ADP_GLC_PYROPHOSPH_3	290	300	ASMGIYVIskH
	PS00808	ADP_GLC_PYROPHOSPH_1	55	74	GGGaGTrLfpLTkrrAkPAV WFqGTAHAV
Potato	PS00809	ADP_GLC_PYROPHOSPH_2	146	154	
(LS)					
Potato	PS00808	ADP_GLC_PYROPHOSPH_1	97	116	GGGaGTrLypLTkkrAkPAV
(SS)	PS00809	ADP_GLC_PYROPHOSPH_2	186	194	WFqGTADAV
	PS00810	ADP_GLC_PYROPHOSPH_3	298	308	ASMGIYVIskD
Arabidopsis	PS00808	ADP_GLC_PYROPHOSPH_1	96	115	GGGaGTrLfpLTkrrAkPAV
(LS)	PS00809	ADP_GLC_PYROPHOSPH_2	186	194	WFqGTADAV
	PS00810	ADP_GLC_PYROPHOSPH_3	301	311	ASMGVYVFkkE
Arabidopsis	PS00808	ADP_GLC_PYROPHOSPH_1	96	115	GGGaGTrLypLTkkrAkPAV
(SS)	PS00809	ADP_GLC_PYROPHOSPH_2	185	193	WFqGTADAV
	PS00810	ADP_GLC_PYROPHOSPH_3	297	307	ASMGIYVVsrD