Characterization and molecular modeling of Inositol 1,3,4 tris phosphate 5/6 kinase-2 from *Glycine max* (L) Merr.: comprehending its evolutionary conservancy at functional level

Ashish Marathe¹ · Veda Krishnan¹ · Mahesh M. Mahajan² · Vinutha Thimmegowda¹ · Anil Dahuja¹ · Monica Jolly¹ · Shelly Praveen¹ · Archana Sachdev¹

Received: 2 August 2017 / Accepted: 26 December 2017 / Published online: 3 January 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Soybean genome encodes a family of four inositol 1,3,4 trisphosphate 5/6 kinases which belong to the ATP-GRASP group of proteins. Inositol 1,3,4 trisphosphate kinase-2 (*GmItpk2*), catalyzing the ATP-dependent phosphorylation of Inositol 1,3,4 trisphosphate (IP3) to Inositol 1,3,4,5 tetra phosphate or Inositol 1,3,4,6 tetra phosphate, is a key enzyme diverting the flux of inositol phosphate pool towards phytate biosynthesis. Although considerable research on characterizing genes involved in phytate biosynthesis is accomplished at genomic and transcript level, characterization of the proteins is yet to be explored. In the present study, we report the isolation and expression of single copy *Itpk2* (948 bp) from *Glycine max* cv Pusa-16 predicted to encode 315 amino acid protein with an isoelectric point of 5.9. Sequence analysis revealed that *GmITPK2* shared highest similarity (80%) with *Phaseolus vulgaris*. The predicted 3D model confirmed 12 α helices and 14 β barrel sheets with ATP-binding site close to β sheet present towards the C-terminus of the protein molecule. Spatio-temporal transcript profiling signified *GmItpk2* to be seed specific, with higher transcript levels in the early stage of seed development. The present study using various molecular and bio-computational tools could, therefore, help in improving our understanding of this key enzyme and prove to be a potential target towards generating low phytate trait in nutritionally rich crop like soybean.

Keywords *GmItpk2* · Low phytic acid · Phylogeny · Real time · Southern · SwissDock

Introduction

Inositol phosphates are imperative compounds for plant metabolism and known to play multifunctional roles in seed germination, sugar transport, auxin signaling, biosynthesis of cyclitols, maintaining hormonal homeostasis (Loewus and Loewus 1983). Developing seeds of legumes like soybean tend to accumulate higher phosphorus, which is more than required for cellular activities. This phosphorus is stored in the form of phytic acid (PA) (myo-inositol

1,2,3,4,5,6 hexa-*kis* phosphate) which is deposited as salts in the vacuole. Being one of the most nutritive crops among agricultural commodities, soybean needs suitable interventions to reduce the inherent phytate levels (~ 4%) for better economic value. The intrinsic nature to form stable complexes with metal ions and proteins has conferred PA an antinutrient property (Pallauf and Rimbach 1997). PA biosynthesis in plants occurs via both “lipid-dependent” and “lipid-independent pathway” of which the lipid-independent pathway predominates during seed development (Stevenson-Paulik et al. 2005; Pandey et al. 2016). The process of PA biosynthesis requires the myo-inositol backbone, which is synthesized from D-Glucose-6-phosphate, catalyzed by myo-inositol-3-phosphate synthase (MIPS) and serves as the primary substrate for both the pathways leading towards phytate biosynthesis. Following the formation of myo-inositol-3-phosphate, sequential phosphorylation reactions via action of various inositol phosphate kinases-myoinositol kinase (*Mik*), Inositol 1,3,4 trisphosphate kinase (*Itpk*), Inositol 1,4,5 trisphosphate 3/6 kinase, (*Ipk2*) and Inositol

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13205-017-1076-z>) contains supplementary material, which is available to authorized users.

✉ Archana Sachdev
arcs_bio@yahoo.com

¹ Division of Biochemistry, ICAR-IARI, New Delhi, India

² Division of Molecular Biology and Biotechnology, ICAR-IARI, New Delhi, India

1,3,4,5,6 penta-*kis* phosphate kinase (*Ipki*) lead to PA synthesis via the lipid-independent pathway and fine-tune the phosphorous flux towards PA accumulation.

In developing countries, high-phytate grain-based diets are feared to exacerbate iron and zinc malnutrition. On the other hand, in developed countries, live stocks primarily fed with grain-based feed are escalating the environmental phosphorus pollution leading to eutrophication (Bohn et al. 2008; Kumari et al. 2014). Being a common problem for both developing and developed countries, humongous interest has surfaced towards reducing the endogenous phytate levels in nutritionally important crops like soybean. Breeding strategies are limited due to lack of natural variation in phytate levels in the available soybean germplasm, and hence genetic manipulation via mutations and metabolic engineering seems to be an alternative approach in developing low phytate soybean with increased mineral bioavailability and thus enhancing the nutritional value.

In this direction, mutation breeding has been successfully applied to generate low phytic acid (*lpa*) mutants in crops like maize (Shi et al. 2003; Shukla et al. 2004), rice (Liu et al. 2007), and *Arabidopsis* (Kim and Tai 2011). Two *lpa* mutants in rice (*Os-lpa*-XS110-1 and *Os-lpa*-XS110-1) were generated by γ -irradiation (Liu et al. 2007), which contained lower PA content ranging from 34 to 64% with respect to the parental lines. In soybean, two *lpa* mutations were isolated and characterized—*Gm-lpa*-TW-1 and *Gm-lpa*-ZC-2 which showed PA reduction of 66.6 and 46.3%, respectively (Yuan et al. 2007). In addition, two soybean *lpa* mutants have been reported; V995089 carrying a mutation in the *MIPS1* gene and CX1834 with a mutation in two multidrug resistant protein genes resulting in 50 and 80% reduction in seed phytate, respectively (Gillman et al. 2009; Maroof et al. 2009). Degradative pathway of phytate was also manipulated by overexpressing phytase in the developing seeds resulting in low phytate levels, which in turn increased the bioavailability of phosphorus (Birch-Pederson et al. 2002). Reverse genetics approach like RNAi-mediated gene silencing has also been employed in various crops like rice, aiming at reducing the inherent seed phytate content (Ali et al. 2013). Applying this technique, the rate-limiting step, catalyzed by *GmMIPS1*, which shifts the metabolic flux from glucose towards phytate, was silenced in soybean that resulted in a drastic reduction of the phytate content up to 94.5% (Nunes et al. 2006). Similarly, the candidate *MRP4* in maize was also silenced by RNAi approach to reduce the phytate content (Shi et al. 2007).

Among the various kinases involved, myo-inositol 1,3,4 tris phosphate 5/6 kinase (ITPK) is a key player and is central to both lipid-dependent and lipid-independent pathway of PA biosynthesis. First discovered in *Arabidopsis*, ITPKs belong to ATP-GRASP family of proteins which catalyse the phosphorylation of myo-inositol 1,3,4 tris phosphate (IP_3)

to myo-inositol 1,3,4, 5/6 tetra phosphate (IP_4). Being conserved from plant kingdom to humans, they are also found in protozoan *Entamoeba histolytica* (Field et al. 2000). The existence of isoforms of ITPKs has been reported in various crops. Six isoforms of ITPK have been reported in rice (Suzuki et al. 2007), four in soybean and wheat (Stiles et al. 2008; Bhatti et al. 2014) and one in maize (Shi et al. 2003). In addition to phosphorylating (IP_3), ITPK family of proteins are also known to phosphorylate the 1-position of Ins (3,4,5,6) P_4 leading to synthesis of Ins (1,3,4,5,6) P_5 .

In spite of the concern towards inositol phosphate signaling and progress in characterizing the genes implicated in PA synthesis, very little is known regarding the structure, regulation and catalysis of the candidate proteins. Till date *E. histolytica* ITPK (*EhITPK*) and Human inositol kinase (*hITPK*) are the only two members of PA synthesis pathway whose 3D structures and molecular models have been elucidated (Miller et al. 2005; Chamberlian et al. 2007). Various physico-chemical properties like size, shape, hydrophilicity and structural features like 3-dimensional configuration, molecular flexibility of a protein determine its functional behaviour in vivo. Physical and enzymatic alterations have been a conventional tool in improving the functionality of a protein and therefore understanding the structural features through various bio-computational tools could provide new avenues to enhance the functionality of a protein at molecular level.

Materials and methods

Plant material

In the present study, *Glycine max* var. Pusa-16, grown in the fields of Division of Genetics, I.A.R.I, New Delhi was used. For temporal expression analysis, developing seeds of various sizes with intact pods were collected at regular intervals starting from 30 days after sowing. To study the spatial variation in expression profiles of *Gmltpk2*, leaves, flowers, stem and roots were collected. Samples were stored in -80°C till further use.

Total RNA isolation, amplification and cloning of *Gmltpk2*

Seeds were sorted into various stages of development according to size (0–4, 4–8, 8–12 and 12–16 mm) based on the length which was measured from the apical to basal end of the seeds of soybean cultivar Pusa-16. Total RNA was extracted from 100 mg of developing seeds using TRI-ZOL method. RNA was quantified spectrophotometrically using Nanodrop 2000 (Thermo Scientific, USA). cDNA was synthesized using RevertAid™ H Minus First Strand

cDNA Synthesis Kit (ThermoScientific, USA). The cDNA was synthesized by 1 µg of total RNA and further used as a template to amplify the full length coding sequence with a set of gene-specific primers (Table 1). Reverse Transcriptase PCR was carried out in a thermocycler in a 25 µL reaction volume containing 1X Taq Buffer, 0.4 Units Phusion High-Fidelity DNA polymerase, 2.5 mM MgCl₂, 0.2 mM dNTP in each tube. The PCR conditions were as follows: 95 °C for 5 min, 35 cycles of 95 °C for 1 min, 60.6 °C for 45 s, 72 °C for 1 min; an additional extension at 72 °C for 10 min. The amplicon was cloned into pGEM[®]-T-Easy vector according to the manufacturer's protocol (Promega, USA). DH5α strain of *E. coli* was transformed with the vector containing the amplicon and recombinant clones were selected on medium containing Ampicillin (50 mg/ml), X-gal (100 µg/ml) and IPTG (0.5 mM). To select the transformed colonies, blue–white method of screening was employed. Plasmid was isolated from the white colonies using Alkaline-Lysis Method (Birnboim and Doly 1979) and screened through restriction analysis using *EcoRI*.

Spatio-temporal transcript profiling of *Gmltpk2*

Developing soybean seeds of various sizes, 0–4, 4–8, 8–12 and 12–16 mm size and tissues from leaves, flowers, seedlings, roots and stem from the plants growing under field conditions were harvested as samples for total RNA extraction and the RNA was isolated using the standard TRIZOL method (Simms et al. 1993). The expression profiling was carried out using both semi-quantitative and quantitative analysis. For semi-quantitative PCR, the primer pairs mentioned in Table 1 were used for amplifying *Gmltpk2*. For quantitative real time PCR, primers were designed using the Oligoanalyzer tool of Integrated DNA Technologies (<http://eu.idtdna.com/calc/analyzer>) (Table 1). 25 µL reaction volume containing 1X Taq Buffer, 0.4 Units Phusion High-Fidelity DNA polymerase, 2.5 mM MgCl₂, 0.2 mM dNTP was set with following PCR conditions: 95 °C for 5 min, 35 cycles of 95 °C for 1 min, 60.6 °C for 45 s, 72 °C for

1 min; an additional extension at 72 °C for 10 min. The qRT-PCR analysis was performed using PikoReal real-time PCR system (Thermo Scientific). To normalize the data, 2^{−ΔΔCt} method (Livak and Schmittgen 2001) was followed using *PEPCo* as the internal control (Table 1). The experiment was carried out in triplicate using DyNAmo ColorFlash SYBR qPCR kit (ThermoScientific, USA). The PCR condition were as follows- 94 °C for 2 min followed by 40 cycles of 94 °C for 30 s, 60 °C for 30 s and 72 °C for 30 s. Data were represented as means and standard errors of the triplicates.

In silico characterization of *GmlTPK2*

Sequence analyses

Sequence analysis of *Gmltpk2* was done against other dicot and monocot *ltpk2* plants using BlastN and BlastP programs of National Centre for Biotechnology Information (NCBI; <https://www.ncbi.nlm.nih.gov>). The amino acid sequences of the dicot and monocot plant species served as templates for phylogenetic studies. MEGA (Molecular Evolutionary Genetic Analysis) version 6 software (<http://megasoftware.net/>) was implemented for constructing the phylogeny tree using the Neighbour Joining method. The physico-chemical properties like amino acid composition, pI, molecular weight, half-life and hydropathy index were determined using ProtParam (<http://web.expasy.org/cgi-bin/protparam/protparam>) (Table 2). Interaction of *GmlTPK2* with other proteins of phytate biosynthetic pathway was predicted using STRING server (<https://string-db.org/>). To analyze the conserved motifs among protein sequences related to *GmlTPK2*, MEME (<http://meme-suite.org>) programme was used.

Structural analysis and homology-based modeling

The 3D structure of the target protein was generated by SWISS MODEL (<https://swissmodel.expasy.org/>) using human inositol phosphate kinase as a template. The predicted structure was further validated via RAMPAGE tool (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>). The secondary structure of the protein was elucidated using PDBsum (<http://www.ebi.ac.uk/pdbsum>). Probability of protein disorder was determined by the PrDOS (Protein disorder prediction server) tool. (<http://prdos.hgc.jp>). The solvent accessibility of the generated protein model was accessed through ASA-view (<http://www.abren.net/asaview/>).

Active site mapping, cleft analysis and molecular docking

The amino acid residues present in the ligand-binding sites were analyzed using FunFold2 server (<http://www.reading.ac.uk/bioinf/FunFOLD/>) and I-TASSER (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>). The cleft analysis to

Table 1 Primers used for standard and real time PCR

Gene	Primer	Amplicon size (in bp)
<i>Gmltpk2</i>	Fp: ATGTCCGAGTCGGAAGTA GCAG	948
	Rp: CTACGCAGTCTTGGAGCGTAC AAT	
<i>Gmltpk2_qRT</i>	Fp: ACGCCAAACAACACAGCATC Rp: CGGGAGGGTCGATGATTACG	180
<i>PepCo_qRT</i>	Fp: CATGCACCAAAGGGTGTITTT Rp: TTTTGTCCGCGAGTCTATCTCTC	200

Table 2 Physico chemical properties of ITPK2 from related monocot and dicot species calculated using ProtParam tool; Isoelectric Point (pI); extinction co-efficient (EC) ($M^{-1}cm^{-1}$); instability index (Ii); aliphatic index (Ai); grand average hydropathy (GRAVY); number of negative residues (– R); number of positive residues (+ R)

Organism	Amino acid residues	Mol. wt. (Dalton)	pI	EC	Ii	Ai	GRAVY	– R	+ R
<i>GmITPK</i>	315	35,013.10	5.90	15,930	46.62	98.63	– 0.169	40	32
<i>PvITPK</i>	317	35,358.41	5.72	22,920	47.10	96.81	– 0.212	42	34
<i>CcITPK</i>	311	34,817.83	6.05	17,420	41.41	96.27	– 0.200	35	29
<i>AiITPK</i>	315	35,085.00	6.11	15,930	46.62	97.71	– 0.207	38	32
<i>AdITPK</i>	352	39,559.40	7.70	20,400	37.60	95.26	– 0.304	38	39
<i>MtITPK</i>	318	35,875.09	5.78	15,930	42.74	94.03	– 0.221	39	31
<i>AtITPK</i>	319	36,219.85	6.03	27,390	51.19	95.2	– 0.218	45	40
<i>GsITPK</i>	166	18,067.71	4.98	7450	42.16	101.51	– 0.033	24	17
<i>TcITPK</i>	310	34,779.94	5.29	18,910	47.39	97.71	– 0.222	42	34
<i>BrITPK</i>	320	36,496.10	6.15	27,390	53.51	91.59	– 0.293	47	44
<i>ZmITPK</i>	355	37,324.35	4.43	5500	34.69	105.10	– 0.138	80	34
<i>JcITPK</i>	335	37,311.80	6.26	17,420	47.29	100.63	– 0.217	39	36
<i>VaITPK</i>	317	35,337.45	5.84	17,420	47.09	95.02	– 0.237	43	36
<i>EhITPK</i>	319	36,480.07	5.95	29,450	35.91	96.80	– 0.239	41	35
<i>hITPK</i>	414	45,621.05	5.78	21,430	37.82	90.22	– 0.155	40	49

Gm, *Glycine max*; *Pv*, *Phaseolus vulgaris*; *Cc*, *Cajanus cajan*; *Ai*, *Arachis iapensis*; *Ad*, *Arachis duranensis*; *Mt*, *Medicago trunculata*; *At*, *Arabidopsis thaliana*; *Gs*, *Glycine soja*; *Tc*, *Theobroma cocoa*; *Br*, *Brassica rapa*; *Zm*, *Zea mays*; *Jc*, *Jatropha curcas*; *Va*, *Vigna angularis*; *Eh*, *Entamoeba histolytica*; *h*, *Homo sapiens*

detect the ligand-binding domains of the protein was done using FTSite Server (<http://ftsites.bu.edu/>). Docking studies were executed to investigate the probable binding modes of the substrates to the active site of *GmITPK2*, for which, PDB file of the modelled *GmITPK2* was imported into SwissDock module (<http://www.swissdock.ch>). The docking results were viewed using UCSF Chimera 1.11rc package (www.cgl.ucsf.edu/chimera).

Molecular characterization of *Gmltpk2*

To determine the native copy number of *Gmltpk2*, Southern hybridization was performed (Southern 1975). Genomic DNA was extracted from the soybean leaves using CTAB method. 10 µg of DNA was separately digested using *EcoRI*, *BamHI* and *HindIII*. The reaction mix was incubated at 37 °C for 16 h. The digested DNA was electrophoresed on 1% agarose gel. The gel was treated with depurination buffer (1% HCl) for 15 min followed by denaturation solution (1.5 M NaCl and 0.5 N NaOH) for 45 min. The gel was then incubated in neutralization buffer (1.5 M NaCl, 0.5 M Tris. HCl, pH 7.4). The DNA was then blotted to nylon membrane by capillary transfer with 10X SSC as the transfer solution. After the transfer was done, the DNA was fixed to the nylon membrane by UV rays in a UV Crosslinker. The membrane was then hybridized with biotin-labeled probe in hybridization solution containing 5X Denhardt Reagent, 6X SSC, 50% Formamide, 0.5% SDS and denatured salmon sperm DNA to a final concentration of 50 µg/mL and incubated at 42 °C for

16 h. Biotin-labeled probes were synthesized using Biotin Decalabel DNA Labeling kit (Thermo Scientific, USA). A 948 bp PCR fragment was purified, eluted and 500 ng eluant was used as a template in this assay. The incorporation of biotin-dUTP was done as per the manufacturer's protocol. In the post hybridization, the membrane was washed twice with 2X SSC, 0.1% SDS for 10 min each at room temperature, followed by two more washes with 0.1X SSC, 0.1% SDS for 20 min each at 65 °C. After washing, the excess washing solution was removed by placing the membrane on a filter paper for a brief period of time. The detection was carried using Biotin Chromogenic Detection kit (Thermo Scientific, USA) as per the manufacturer's protocol.

Results and discussion

Cloning of full length *Gmltpk2* and sequence analysis

Four isoforms of *Itpks* (*Gmltpk1-4*) have been reported in soybean of which *Gmltpk3* was reported to show seed-specific expression (Stiles et al. 2008). Considering the hint on role of other *Itpks* like 1 and 2 from expression studies on developing soybean seeds (data not shown), *Gmltpk1*, in Pusa-16 has been cloned and characterized previously (Krishnan et al. 2015). To explore the structural insights of *Itpk2*, in the present study, a full-length *Gmltpk2* coding sequence was isolated from Pusa-16

cultivar of soybean using gene-specific primers. The *GmItpk2* cDNA contained a 948 bp long open reading frame and the amplicon was viewed on 1% agarose gel stained with ethidium bromide (Fig. 1a). Cloning was confirmed through restriction analysis showing a release of 948 bp fragment (Fig. 1b). The sequence was submitted to NCBI GenBank database (Accession No. KJ579138.1). The cDNA encoded a protein of 315 amino acids with a molecular weight of 35 kDa as calculated by ProtParam tool of ExPASy package. The amino acid sequence of *GmITPK2* was deduced from the nucleotide sequence using Softberry tools (<http://www.softberry.com>). The predicted protein sequence revealed 80% relatedness with hypothetical protein from *Phaseolus vulgaris* (Accession no. XP_007155434.1). Using the Neighbour Joining method, a phylogenetic tree was generated to evaluate the evolutionary relatedness of *GmITPK2* with other selective monocot and dicot members of plant kingdom (Fig. 2). The optimized search of the ITPK2 proteins from various monocot and dicot species showed 3 different conserved motifs (Fig. S1). All the sequences had 3 motifs, but in case of *GsITPK*, the second motif was absent (Fig. S2). The cytosolic localization of *GmITPK2* protein was predicted by CELLO tool with a reliability index of 2.976. The presence of destabilising amino acids in the *N*-terminus, MSESEVAGQ imparts an unstable nature to the protein which is justified by its instability index of 46.62 with a $T_{1/2}$ of less than 5 h (Rogers et al. 1986). Four disordered regions were predicted in the protein sequence, of which the longest disordered region was found between Ser²²⁸ to Asp²⁴¹ comprising 14 amino acid residues (Fig. 3). ExPASy ProtParam was employed to analyse the physico-chemical properties of the selected ITPKs from different dicot and monocot species (Table 2). The analysis revealed Leucine as the most abundant amino acid for *GmITPK2*, accounting for about 11%, while

Tryptophan, Methionine and Cysteine were the least abundant. ProtParam analysis computed the isoelectric point of the protein to be 5.90, and is therefore expected to precipitate in acidic buffer system. The extinction coefficient was $15,930 \text{ M}^{-1}\text{cm}^{-1}$ considering all the cysteine reduced in a reduced form since the protein was cytosolic in nature. A low GRAVY of -0.169 indicates the hydrophilic nature of the protein which is attributed to charged amino acid residues present in the protein sequence (40 negatively charged and 32 positively charged). To validate GRAVY index to the hydrophilic nature of *GmITPK2*, surface characterization of the hydrophobicity was done to predict the hydrophilic domains present in the protein. Chimera 1.11rc analysis revealed that a greater portion of the protein contained polar amino acids, which are hydrophilic in nature (Fig. S3). This result also confirms the cytosolic nature of *GmITPK2*, and its interaction with various other proteins involved in phytate biosynthesis, due to its polar nature. Thermostability of proteins is generally determined by its Aliphatic index (Ai) (Gupta et al. 2012). The Aliphatic index of the ITPK proteins in the present study ranged from 86.6 (*Cajanus cajan*) to 105.1 (*Zea mays*). High Ai of these proteins indicates a higher thermostability of various ITPK proteins and is predicted to play a role in response to various biotic and abiotic stresses. Recent reports in crops like sweet potato, *Arabidopsis* and rice suggest the role of various enzymes of phytate biosynthetic pathway in providing tolerance mechanism to stress (Zhai et al. 2016; Yang et al. 2008; Niu et al. 2008). This attribute can possibly be explored in studies pertaining to cell signaling under biotic and abiotic stress. Solvent accessibility of *GmITPK2* protein was done to deduce the composition of amino acids on the surface and the innercore using ASA view (Fig. S4). Since phytate biosynthesis is restricted to the cytosol, the protein interaction studies were performed by STRING

Fig. 1 **a** PCR amplification of *GmItpk2* gene. Lane nos. 1–3 show amplicons (948 bp); M- Molecular weight marker (1 kb DNA ladder). **b** Confirmation of recombinant *E. coli* clones by restriction digestion with *EcoRI* restriction enzyme. Lane nos. 1–3 represent positive recombinant clones by releasing of a 948 bp DNA fragment; M- Molecular weight marker (1 kb DNA ladder)

Fig. 4 Spatio-temporal expression analysis of GmItpk2 through quantitative PCR

reached maturity. The transcript levels were however very low in the other tissues (less than 0.5 fold) thereby confirming its seed-specific nature (Fig. 4). Expression during early stages unambiguously indicates the role of *GmItpk2* in inositol phosphates accumulation which are also required for phospholipids biosynthesis and membrane development besides being diverted towards PA synthesis. The results were in confirmation with the expression studies of *Itpka* in *Phaseolus vulagris* (*PvItpk*) wherein expression of *PvItpka* was found higher at 12–14 DAF (days after flowering) with a reduction in expression at 20–24 DAF (Fileppi et al. 2010). Coelho et al. (2005) also reported accumulation of lower inositol phosphates—IP₃, IP₄ and IP₅ during seed development from 16 DAF to 26 DAF in common bean, while in soybean, the accumulation of IP₆ was observed 30 DAF (Basak et al. 2017). Among the three inositol phosphates, IP₃ levels were found to be maximum in the early stage. Expression of *Itpk* during early stages results in increased inositol phosphate pool which is required during the cell division and signaling in developing seeds. The increased inositol phosphate levels during early stages could be attributed to the kinase activity of *GmItpk2*, converting Ins 1,3,4 tris phosphate into Ins 1,3,4,5/6 tetra phosphate or Ins 3,4,5,6 tetra phosphate into Ins 1,3,4,5,6 penta phosphate, as well as the phosphatase activity leading to synthesis of Ins 3,4,5,6 tetra phosphate (Stiles et al. 2008).

Homology modeling and structural characterization

Since no plant inositol kinase template was available, the homology model of *GmITPK2* was generated employing SWISSMODEL server using Inositol- tetrakisphosphate- 1 kinase (*Homo sapiens*) as the template (Fig. 5). Five models were generated in total, of which the best model chosen was based on high resolution and per cent coverage.

Fig. 5 3D model of GmITPK2 generated via homology-based modeling using SWISS MODEL depicting various secondary structures— α helices, β pleated sheets and random coils

The sequence identity score was 28.95 with a resolution of 2.01 Å. The 3D model generated was further validated using RAMPAGE program (Fig. 6) and submitted to Protein Model Database (PM0079933). The torsion angles, ψ and ϕ were examined to access the reliability of the protein model. The results obtained in the validation, 97.1% of the amino acid residues were found in the most favoured region, while 2.2 and 0.6% of the amino acid residues were found in the allowed region and the outlier region, respectively. The secondary structure generated with PDBsum predicted a total of 12 α helices (34.29%), 14 β pleated sheets (23.17%) (Fig. 7). The secondary structure of *EhITPK* comprised of a similar topology with 11 α helices and 14 β pleated sheets (Miller et al. 2005). Secondary structure of Human inositol 1,3,4 trisphosphate 5/6 kinase (ITPK 5/6 kinase) predicted by Chamberlian et al. (2007) comprised of 12 α helices and 14 β pleated sheets. The functional activity of *GmITPK2*, *EhITPK* and Human ITPK 5/6 kinase is evolutionarily well conserved owing to their topological similarity, since maintaining the core structure is a prerequisite to preserve the protein function.

Active site mapping and molecular docking

Active site mapping for determining the residues involved in binding to the above mention ligands was done using I-TASSER and Funfold server. The residues Arg⁹⁷, Glu¹⁴⁶, Glu¹⁴⁸, Ala¹⁶¹, Gln¹⁸⁰, Glu¹⁸¹, Phe¹⁸², Val¹⁸³, Val¹⁸⁹, Ile²⁷², Ile²⁸⁴ bind to ATP (3r5xC) with highest C-score of 0.75 (Fig. 8a), whereas Asp²⁸⁵ and Asn²⁸⁷ were present in Mg⁺² (4i50F)-binding site with a score of 0.1. The ATP-binding

(a)

(b)

Fig. 6 Validation of 3D predicted structure using **a** RAMPAGE and **b** PDBSum

site is located in the β sheet close to the C-terminus (amino acid residue 108–314) containing the GRASP domain predicted by FTsite server (Fig. 8b). The residues Lys²¹, Ser²⁴, Lys⁶², Asp¹⁵³, Gly¹⁵⁴, His¹⁵⁹, Lys¹⁹¹, Tyr¹⁹³, Asp²⁷⁰, Asp²⁸⁵, Asn²⁸⁷, Pro²⁹⁰ and Phe²⁹¹ are predicted to bind inositol tetra phosphate (IP₄). Kinetic studies for substrate specificity by Stiles et al. (2008) also revealed inositol tetra phosphate (IP₄) as a preferred substrate over Inositol tris

Fig. 7 Schematic representation of the secondary structure prediction of GmITPK2 using PDBSum tool. Arrows (Pink) indicating the β pleated sheets and Barrels (Red) indicating the α Helices

Fig. 8 **a** Amino acid residues present at the catalytic site predicted by FunFold Server 2. Ligand-binding residues shown in blue. **b** Cleft Analysis using FTSite server predicting the ATP-binding site between two β pleated sheets of GmITPK2. **c** Representative binding mode of ATP at the active site of GmITPK2 subsequent to docking simulation using Swiss-Dock

phosphate (IP₃). The kinetic studies using labeled Inositol 1,3,4, tris phosphate by Phillippy (1998) showed Inositol 1,3,4,5 tetra phosphate as the preferred substrate which was converted into Ins 1,3,4,5,6 penta phosphate. Earlier

report by Miller et al. (2005) in the crystal structure of *EhITPK* has shown the presence of Pro¹⁰⁹, Ile¹³⁴, Tyr¹⁷⁰, Ile¹⁷¹, Leu¹⁹⁵, Phe²⁰⁸, and Ile²⁷⁷ lining the nucleotide-binding pocket whereas Asp²⁸⁹ and Asn²⁹¹ were predicted to bind to Mg²⁺. Lys¹⁷⁹, His¹⁴⁷, Arg¹⁹² and Ser²⁹⁵ were predicted to be present in the IP-binding site. A similar pattern of amino acid residues lining the nucleotide-binding and inositol phosphate-binding site was observed in HumanITPK1 (hITPK), with Asp²⁹⁵ and Asn²⁹⁷ binding to ATP-Mg²⁺ while His and Lys were present in the inositol phosphate-binding site. The presence of His in the substrate-binding site among the ITPK proteins is predicted to play a role in intersubstrate phosphate transfer (Chamberlian et al. 2007). Although *GmITPK2* shares a sequence similarity as low as 25% with that of *EhITPK* and 31% with hITPK, the presence of the similar amino acid residues in the catalytic domain indicates the evolutionary importance and conservancy of ITPK proteins during the course of evolution. The presence of conserved amino acid residues in the nucleotide-binding and inositol-binding domain of *GmITPK2* indicates the potential for protein engineering and altering the protein activity via targeted mutagenesis. To spot the potential ligand-binding sites, ATP, which was found to be a preferred ligand, was docked with *GmITPK2*. The docking result viewed with UCSF Chimera 1.11rc (Fig. 8c). Soybean, being a target for functional genomics, where nutritional traits are maneuvered by γ -irradiation, such in silico protein models can prove beneficial in predicting their metabolic key role prior to protein engineering (Krishnan et al. 2016, 2018).

Genomic organization of *Gmltpk2*

The results of Southern hybridization revealed the presence of a single copy of *Gmltpk2*, resulting from a single hybridising band, in soybean genome (Fig. 9). Fileppi et al. (2010) provided similar reports suggesting the presence of a single copy number of genes involved in phytate biosynthesis in *P. vulgaris*, except for myo-inositol phosphate synthase (*MIPS*). The Southern blotting results were confirmed through BLAT tool (<https://legumeinfo.org/blat>) which showed the presence of a single copy with a match score of 100.00. The presence of a single copy in the genome would, therefore, prove beneficial in terms of minimising the functional redundancy which could possibly decrease the consequence of gene silencing. The genomic sequence was further characterized using Phytozome 12 tool (<https://phytozome.jgi.doe.gov>). On the basis of alignment between the genomic and cDNA sequences, the *Gmltpk2* gene on chromosome number 17 entailed a continuous reading frame and was found to be intronless (Fig. S6).

Fig. 9 Southern blot analysis of *Gmltpk2* gene confirming single copy presence in the soybean genome after digestion with BamHI (Lane 2), HindIII (Lane 3) and EcoRI (Lane 4) restriction enzymes. Lane 1 showing positive control

Conclusion

The complete coding sequence of *Gmltpk2* from *Glycine max* var. Pusa 16 was cloned and characterized and its expression analysis revealed the abundance of its transcript levels during early stage of seed growth and development, implicating its key role in seed inositol phosphate biosynthesis, required for cell signaling and phytate biosynthesis. Bio-computational tools were employed to comprehend the structural attributes like secondary structure, physico-chemical properties and the amino acid residues participating in the catalytic activity. 3D model of the protein was generated via homology-based modeling which can help in understanding the structural and functional characteristics of the protein. From the in silico studies, it is clear that the functional character of *GmITPK2*, *EhITPK* and hITPK as 1,3,4 tris phosphate 5/6 kinase is well conserved during the evolution, which is attributed to similarity in their topology, which determines the protein function. The present study is the first report on homology modeling of *GmITPK2* and its 3D model has been submitted to Protein Model Database (PM0079933). This model can serve as a template for characterizing ITPKs from plant species since no templates are available for the same in the protein model database.

Acknowledgements The authors are thankful to DST-INSPIRE (IF130323), Govt. of India, for providing the financial assistance.

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest regarding the publication of this article.

References

- Ali N, Paul S, Gayen D, Sarkar SN, Datta K, Datta SK (2013) Development of low phytate rice by RNAi mediated seed-specific silencing of inositol 1,3,4,5,6-pentakisphosphate 2-kinase gene (*IPK1*). *PLoS One* 8(7):1–12
- Basak N, Krishnan V, Pandey V, Punjabi M, Hada A, Marathe A, Jolly M, Sachdev A (2017) Molecular characterization of inositol pentakisphosphate 2-kinase (*GmIPK1*) from soybean and its expression pattern in the developing seeds. *Indian J Genet* 77(3):371–378
- Bhatti KK, Aggarwal S, Sharma S, Mantri S, Singh SP, Bhalla S, Kaur J, Tiwari S, Roy JK, Tuli R, Pandey AK (2014) Differential expression of structural genes for the late phase of phytic acid biosynthesis in developing seeds of wheat (*Triticum aestivum* L.). *Plant Sci* 224:74–85
- Birch-Pederson H, Sorensen LD, Holm PB (2002) Engineering crop plants: getting a handle on phosphate. *Trends Plant Sci* 7:118–125
- Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* 7(6):1513–1523
- Bohn L, Meyer AS, Rasmussen KS (2008) Phytate: impact on environment and human nutrition. A challenge for molecular breeding. *J Zhejiang Univ Sci B* 9(3):165–191
- Chamberlian PP, Qian X, Stiles AR, Cho J, Jones DH, Lesley SA, Grabau EA, Shears SB, Spraggon G (2007) Integration of inositol phosphate signalling pathways via Human ITPK1. *J Biol Chem* 281:28117–28125
- Coelho CMM, Tsai SM, Vitorello VA (2005) Dynamics of inositol phosphate pools (tris-, tetrakis- and pentakisphosphate) in relation to the rate of phytate synthesis during seed development in common bean (*Phaseolus vulgaris*). *J Plant Physiol* 162(1):1–9
- Field J, Wilson MP, Mai Z, Majerus PW, Samuelson J (2000) An *Entamoeba histolytica* Inositol 1, 3, 4-trisphosphate 5/6-kinase has a novel 3-kinase activity. *Mol Biochem Parasitol* 108:119–123
- Fileppi M, Galasso I, Tagliabue G, Daminati MG, Campion B, Doria E, Sparvoli F (2010) Characterisation of structural genes involved in phytic acid biosynthesis in common bean (*Phaseolus vulgaris* L.). *Mol Breeding* 25:453–470
- Gillman JD, Pantalone VR, Bilyeu K (2009) The low phytic acid phenotype in soybean line CX1834 is due to mutations in two homologs of the maize low phytic acid gene. *Plant Genome* 2(2):179. <https://doi.org/10.3835/plantgenome2008.03.0013>
- Gupta SK, Rai AK, Kanwar SS, Sharma TR (2012) Comparative analysis of zinc finger proteins 553 involved in plant disease resistance. *PLoS One* 7(8):e42578. <https://doi.org/10.1371/journal.pone.0042578>
- Kim S, Tai T (2011) Identification of genes necessary for wild-type levels of seed phytic acid in *Arabidopsis thaliana* using a reverse genetics approach. *Mol Genet Genom* 286:119–133
- Krishnan V, Jain P, Vinutha T, Hada A, Manickavasagam M, Ganapathi A, Rai RD, Sachdev A (2015) Molecular modeling and in silico characterization of *Glycine max* Inositol (1,3,4) tris 5/6 kinase-1 (*Gmitpk1*)- a potential candidate gene for developing low phytate soybean transgenics. *POJ* 8(5):381–391
- Krishnan V, Singh A, Vinutha T, Singh B, Dahuja A, Rai RD, Sachdev Archana (2016) Low gamma irradiation effects on protein profile, solubility, oxidation, scavenger ability and bio-availability of essential minerals in black and yellow Indian soybean (*Glycine max* L.) varieties. *J Radioanal Nucl Chem* 307(1):49–57
- Krishnan V, Gothwal S, Dahuja A, Vinutha T, Singh B, Jolly M, Praveen S, Sachdev A (2018) Enhanced nutraceutical potential of gamma irradiated black soybean extracts. *Food Chem* 245:246–253
- Kumari S, Krishnan V, Jolly M, Sachdev A (2014) In vivo bioavailability of essential minerals and phytase during soaking and germination in soybean (*Glycine max* L.). *Aust J Crop Sci* 8(8):1168–1174
- Liu QL, Xu XH, Ren XL, Fu HW, Wu DX, Shu QY (2007) Generation and characterization of low phytic acid germplasm in rice (*Oryza sativa* L.). *Theor Appl Genet* 114:803–814
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2^{(-Delta Delta C(T))} Method. *Methods* 25:402–408
- Loewus FA, Loewus MW (1983) myo-Inositol: its biosynthesis and metabolism. *Ann Rev Plant Physiol* 34:137–161
- Maroof MAS, Glover NM, Biyashev RM, Buss GR, Grabau EA (2009) Genetic basis of the low-phytate trait in the soybean line CX1834. *Crop Sci* 49(1):69
- Miller GJ, Wislon MP, Majerus PW, Hurley JH (2005) Specificity determinants in inositol polyphosphate synthesis: crystal structure of inositol 1,3,4-trisphosphate 5/6-kinase. *Mol Cell* 18:201–212
- Niu X, Chen Q, Wang X (2008) *OsITL1* gene encoding an inositol 1,3,4-trisphosphate 5/6-kinase is a negative regulator of osmotic stress signalling. *Biotechnol Lett* 30:1687–1692
- Nunes ACS, Vianna GR, Cuneo F, Rech EL, Aragao FLJ (2006) RNAi-mediated silencing of the myo-inositol-1-phosphate synthase gene (*GmMIPSI*) in transgenic soybean inhibited seed development and reduced phytate content. *Planta* 224:125–132
- Pallauf J, Rimbach G (1997) Nutritional significance of phytic acid and phytase. *Arch Anim Nutr* 50:301–319
- Pandey V, Krishnan V, Basak N, Hada A, Punjabi M, Jolly M, Lal SK, Singh SB, Sachdev A (2016) Phytic acid dynamics during seed development and its composition in yellow and black Indian soybean (*Glycine max* L.) genotypes through a modified extraction and HPLC method. *J Plant Biochem Biot* 25:367–374
- Phillippy BQ (1998) Identification of inositol 1,3,4 trisphosphate 5-kinase and inositol 1,3,4,5-tetra kishphosphate-6 kinase in immature soybean seeds. *Plant Physiol* 116:291–297
- Rogers S, Wells R, Rechsteiner M (1986) Amino acid sequences common to rapidly degraded 621 proteins: the PEST hypothesis. *Science* 234:364–368
- Shi J, Wang H, Wu Y, Hazebroek J, Meeley R, Ertl D (2003) The maize low-phytic acid mutant *Ipa2* is caused by mutation in an inositol phosphate kinase gene. *Plant Physiol* 131:507–515
- Shi J, Wang H, Schellin K, Li B, Faller M, Stoop JM, Meeley RB, Ertl DS, Ranch JP, Glassman K (2007) Embryo-specific silencing of a transporter reduces phytic acid content of maize and soybean seeds. *Nat Biotechnol* 25:930–937
- Shukla S, Van Toai T, Pratt R (2004) Expression and nucleotide sequence of an ins (3) p-1 synthase gene associated with low-phytate kernels in maize (*Zea mays* L.). *J Agric Food Chem* 52:4565–4570
- Simms D, Cizdziel PE, Chomczynski P (1993) TRIzol: a new reagent for optimal single-step isolation of RNA. *Focus* 15(4):532–535
- Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98:503–517
- Stevenson-Paulik J, Bastidas RJ, Chiou ST, Frye RA, York JD (2005) Generation of phytate free seeds in *Arabidopsis* through disruption of inositol polyphosphate kinases. *Proc Natl Acad Sci USA* 102:12612–12617
- Stiles AR, Qian X, Shears SB, Grabau EA (2008) Metabolic and signalling properties of an *Itpk* gene family in *Glycine max*. *FEBS Lett* 582:1853–1858
- Suzuki M, Tanaka K, Kuwano M, Yoshida K (2007) Expression pattern of inositol phosphate-related enzymes in rice (*Oryza sativa* L.): implications for the phytic acid biosynthetic pathway. *Gene* 405:55–64

- Yang L, Tang R, Zhu J, Liu H, Mueller-Roeber B, Xia H, Zhang H (2008) Enhancement of stress tolerance in transgenic tobacco plants constitutively expressing *AtIPK2 β* , an inositol polyphosphate 6/3-kinase from *Arabidopsis thaliana*. *Plant Mol Biol* 66:329–343
- Yuan FJ, Zhao HJ, Ren XL, Zhu SL, Fu XJ, Shu QY (2007) Generation and characterization of two novel low phytate mutations in soybean (*Glycine max* L. Merr.). *Theor Appl Genet* 115:945–957
- Zhai H, Wang F, Si Z, Huo J, Xing L, An Y, He S, Liu Q (2016) A *myo*-inositol-1-phosphate synthase gene, *IbMIPS1*, enhances salt and drought tolerance and stem nematode resistance in transgenic sweet potato. *Plant Biotechnol J* 14:592–602