RNAi-INDUCED TRANSCRIPTIONAL SILENCING OF *INOSITOL POLYPHOSPHATE 6-/3-/5-KINASE* (*IPK2*) GENE IN SOYBEAN SEEDS TO GENERATE LOW PHYTATE LINES

A thesis submitted to
Delhi Technological University
for the award of the degree

of

DOCTOR OF PHILOSOPHY

in

BIOTECHNOLOGY

by

MANSI PUNJABI

DEPARTMENT OF BIOTECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
DELHI, INDIA

MAY 2019

DECLARATION

I, Mansi Punjabi, hereby declare that the research work embodied in this Ph.D. thesis is my own bonafide work and neither any part or the whole of this thesis has been submitted for the award of any other degree.

I further certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright and that any excerpts, techniques, thoughts, phrases, or any other matter from the work belonging to other people cited in my thesis are dully acknowledged. I have not wilfully lifted data, results, etc. from the published and unpublished work of others and included them in the Ph.D. thesis as my own work. All the information in this thesis has been furnished in accordance with academic rules and ethical conduct.

Date: 07/05/2019

Place: Delhi Technological University

Mansi Punjabi
Reg. No. 2k13/PhDBT/04

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Govt. of NCT of Delhi
Delhi-110042, India

CERTIFICATE

This is to certify that the research work embodied in this Ph.D. thesis entitled **"*RNAi induced transcriptional silencing of inositol polyphosphate 6-/3-/5-kinase (IPK2) gene in soybean seeds to generate low phytate lines*"** submitted by **Ms. Mansi Punjabi** (Reg No. 2k13/PhDBT/04) to Delhi Technological University (formerly DCE) is a record of her original work carried out under my supervision. It is further certified that the work has not been submitted to any other university or institution for the award of any other degree.

Dr. Navneeta Bharadvaja
(Supervisor)
Assistant Professor
Department of Biotechnology
Delhi Technological University

Dr. Archana Sachdev
(Co-supervisor)
Principal Scientist
Division of Biochemistry
ICAR-Indian Agricultural Research Institute

Prof. Jai Gopal Sharma
Head
Department of Biotechnology
Delhi Technological University

Acknowledgement

The great American author John C. Maxwell once said, "Nothing of significance was ever achieved by an individual acting alone". This thesis is the outcome of a fortunate journey I had undertaken under the unwavering support and encouragement of various people including my family and friends. I owe my indebtedness to all these people, if it weren't for them, this dissertation would not have been possible.

Firstly, I would like to express my deepest gratitude to my supervisors Dr. Navneeta Bharadvaja and Dr. Archana Sachdev for their collegial guidance and subject matter expertise to address the concerns raised in my thesis. I am thankful to them for giving me the freedom to enthusiastically explore on my own, and instilling in me the qualities of being a good scientist. It has been a greatly enriching experience for me to work under their mentorship. Besides receiving academic help, their friendship has been invaluable to me on a personal level.

I am deeply indebted to the revered Vice-Chancellor Prof. Yogesh Singh, Head of the Department Prof. Jai Gopal Sharma, and other respected faculty members of Delhi Technological University for their academic support and valuable advices throughout my Ph.D. tenure.

I am deeply indebted and would like to express my sincere thanks to Dr. Shelly Praveen, Head, Division of Biochemistry and Dr. Archana Sachdev, Principal Scientist, Division of Biochemistry, Indian Agricultural Research Institute, for their extensive research support by providing generous access to the necessary infrastructure and research tools at their Division.

I would also like to take this opportunity to thank my student research committee members, Prof. Jai Gopal, Head, Department of Biotechnology, Prof. S.K. Garg, Pro Vice Chancellor, DTU, Dr. Anil Dahuja, Principal Scientist, Division of Biochemistry IARI and Prof. Sunil K. Khare, Department of Chemistry, IIT Delhi for critically analyzing my work and giving their constructive advices.

I would like to acknowledge the valuable inputs of faculty members and technical staff at Division of Biochemistry, IARI particularly Mrs. Veda Krishnan, Scientist, Dr. Sweta Kumari, Scientist, and Dr. Monica Jolly, Chief Technical Officer, that helped in shaping my final thesis. I will always be indebted to them for their patience and unfaltering support.

This acknowledgement would be incomplete without mentioning my research lab colleagues, seniors and batch mates. I am grateful to all the members of my group, Ashish Marathe, Nabaneeta Basak, Alkesh Hada, Amit and Vanita Pandey for their friendship and support. I am sincerely thankful to my senior Mrs. Nupur Jauhari for her guidance, advices and constant moral support. I would like to offer my heartfelt thanks to Abhishek Saini for being a true friend and offering his incessant support. I deeply respect and thank Pano Aunty and Naveen for rendering their professional as well as personal support.

I am also thankful to Mr. Chhail Bihari, Mr. Jitendra Singh, Ms. Saumya Maurice and other technical staff members at DTU for their kind support throughout the course of my research.

I would especially like to express my profound gratitude to my husband, Piyush Bathla, for his dedicated support and patience at all times. He helped me survive the

stress from all these years and encouraged me to pursue my academic goals. His unflinching support and selfless care will remain in my heart forever.

I would like to sincerely thank my loving family, my mother-in-law, Mrs. Promila Bathla, father-in-law, Col. P.K. Bathla and brother-in-law, Aayush Bathla. Without their patience and motivation, I would not have been able to complete the momentous task of this thesis.

Above all, I want to thank my parents, Mrs. Maria Elena Punjabi and Late Sh. C.P. Punjabi for being my ultimate teachers, friends and advisers. Their boundless love and support have always provided me strength. Their valuable lessons taught me the true importance of life. I fail for words, to describe everything they have done for me throughout the years. It is more than a blessing to have them as my parents.

Date: *07/05/2019* ***Mansi Punjabi***

Abstract

Soybean is one of the leading oilseed crops in the world and is showing a remarkable surge in its utilization in formulating animal feeds and supplements. Its dietary consumption, however, is incongruent with its existing industrial demand due to the presence of anti-nutritional factors in sufficiently large amounts. Phytic acid, in particular, raises concern as it causes a concomitant loss of indigestible complexed minerals and charged proteins in the waste and results in reduced mineral bioavailability in both livestock and humans. Reducing the seed phytate level thus seems indispensable to overcome the nutritional menace associated with soy grain consumption.

To conceive our objective, we designed and expressed an *inositol poly-phosphate 6-/3-/5-kinase* gene-specific intron-containing self-complementary hairpin RNAi construct in the seeds of Pusa-16 soybean cultivar. *Inositol polyphosphate 6-/3-/5-kinase* is a multifunctional kinase which catalyzes the formation of inositol penta*kis*phosphate, the immediate substrate of PA biosynthesis. Due to a broad range of substrate specificity, it plays a pivotal role in regulating the cellular phytic acid turnover and is therefore conceived to be the most appropriate target for effective seed phytate reduction.

We subsequently conducted a genotypic, phenotypic, and biochemical analysis of the developed putative transgenic populations and found low phytic acid levels, moderate accumulation of inorganic phosphate and elevated mineral content in some lines. These low phytic acid lines did not show any reduction in seedling emergence and displayed an overall good agronomic performance, thus, proving to be a successful attempt.

Contents

List of Figures

List of Tables

List of Abbreviations & Symbols

Å	Angstrom
ABC	ATP-binding cassette
AGO	Argonaute
ANFs	Anti-nutritional factors
B5	Gamborg's Medium
BAP	6-Benzylaminopurine
bp	Base pairs
BSA	Bovine serum albumin
$CaCl_2$	Calcium chloride
CCM	Co-cultivation medium
cDNA	Complementary DNA
CDS	Coding sequence
DDW	Double distilled water
DEPC	Diethyl pyrocarbonate
dNTP	Deoxynucleotide triphosphate
dsRNA	Double stranded RNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EtBr	Ethidium bromide
g	Gram
GA_3	Gibberellic acid
GM	Germination medium
GmIPK2	Glycine max inositol polyphosphate 6-/3-/5-kinase
H_2SO_4	Sulphuric acid
HCl	Hydrochloric acid
hr	Hour(s)
IAA	Indole-3-acetic acid
IBA	Indole-3-butyric acid
ihp	Intron-containing self-complementary hairpin

IMP	Inositol monophosphatase
IPK1	Inositol polyphosphate 2-kinase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITPK	Inositol 1,3,4-trisphosphate 5-/6-kinase
Kb	Kilobase pairs
LA	Luria-Bertani agar
LB	Luria-Bertani broth
lpa	Low phytic acid
MD	Molecular dynamics
mg	Milligram
$MgCl_2$	Magnesium chloride
MIK	*myo*-inositol kinase
min	Minute(s)
MIPS	*myo*-inositol 3-phosphate synthase
mM	Millimolar
MS	Murashige & Skoog medium
MSA	Multiple sequence alignment
NFW	Nuclease free water
nm	Nanometre
ns	Nanosecond
OD	Optical density
P	Phosphorus
PA	Phytic acid
PAT	phosphinothricin acetyl transferase
PDCAAS	Protein digestibility-corrected amino acid score
PEPCo	Phosphoenolpyruvate carboxylase
Pi	Inorganic phosphorus
PTGS	Post-transcriptional gene silencing
qRT-PCR	Quantitative real-time polymerase chain reaction
RISC	RNA-induced silencing complex
RM	Rooting medium
RMSDs	Root mean square deviations

RMSFs	Root mean square fluctuations
RNAi	RNA interference
RNase	Ribonuclease
RP-HPLC	Reversed-phase high-performance liquid chromatography
rpm	Revolutions per minute
RT	Reverse transcriptase, Room temperature
RT-PCR	Reverse transcription polymerase chain reaction
SD	Standard deviation
SDS	Sodium dodecyl sulfate
sec	Second(s)
SEM	Shoot elongation medium
SIM	Shoot induction medium
siRNA	Small interfering RNA
sRNA	Small RNA
SSC	Saline-sodium citrate
T-DNA	Transfer DNA
TBE	Tris-Borate-EDTA
Tbr	Thermus brockianus
TCA	Trichloroacetic acid
UTR	Untranslated region
UV	Ultra Violet
v/v	Volume by volume
w/v	Weight by volume
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
µg	Microgram
µL	Microliter
µM	Micromolar
µm	Micrometer

Chapter 1

Introduction

1.1 General Introduction

The soybean crop, since the twentieth century, has achieved a global status owing to its superior quality protein and oil rich in healthy fats. Popularly known as the ‘golden bean’ or ‘miracle bean’, soybean is an exceptionally nutritious, low glycemic index crop which provides all the essential factors crucial to support human growth and development. Soybean, as an industry, is showing rapid expansion globally. Essential to livestock feeding, soy has become an integral part of our diet as well. Scientists around the world are learning new methods to extract a variety of products from soybeans for human food, animal feed, and industrial applications. Annually, around 85% of the global soybean produce is turned into oil and meal [1]. Majority of the meal (~98%) thus derived is converted to animal feed. The use of soy meal is popular in livestock feed due to its year-round availability, low-cost and consistent high-quality nutrition. The remaining meal is used to prepare soy flour and products including tofu, soy milk, soy sauce, textured vegetable protein, soy protein isolates and concentrates and fermented foods such as natto, tempeh, etc. Soybean holds the first rank amongst oilseed crops in its production, with a world produce of 333.219 million metric tons during the year 2016-2017. Most of its oil fraction is used directly for cooking and in the food industry for manufacturing products such as vegetable shortening, dressings, sauces, mayonnaise, etc. The remaining oil fraction is used for resins, pharmaceuticals, cosmetics, biodiesel and more. According to World Wildlife Fund, soy is expected to fulfil about 10% of EU’s biofuel demand by 2020.

Despite the facts presented above, the true potential of soy to fulfil nutritional requirements in humans and non-ruminant livestock has not yet been explored as they are unable to digest a few compounds accumulated throughout the development and maturation of its seed [2]. In this process, developing soybean seeds accumulate a principal phosphorus (P) storage compound, phytic acid (PA) which accounts for approximately two-thirds of total seed P. Due to its chemical structure, PA precipitates in protein storage bodies in the form of mineral salts especially with iron (Fe), zinc (Zn), magnesium (Mg), potassium (K), calcium (Ca), manganese (Mn), etc. to form phytate salts. During germination, these phytate reserves are hydrolyzed by phytase enzymes to remobilize the stored P and other minerals in order to support seedling growth. Paradoxically, the function that makes PA valuable to the plants constitute antinutritional character for non-ruminants that feed on the grains enriched with it. Monogastrics due to a lack of endogenous phytase in their gastrointestinal tract, fail to process these phytates and thus compromise their mineral bioavailability [3-6]. To compensate for their nutritional P requirement, a common practice is to supplement the feeds with nutrient P or phytase [7]. However, a very small amount of it is effectively absorbed by the animals. On the other hand, the phytate that remains undigested is eliminated in the animal waste which is reintroduced to the agricultural fields in the form of manure [8]. P and other nutrients in runoff from these fields accelerate eutrophication of surface waters and contribute significantly to water pollution [9, 10]. Thus, besides having a negative impact on nutrition, PA also results in serious environmental implications. These reasons, therefore, developed a significant interest of the scientific community to alter phytate production in the soybean crop.

Different strategies have been devised to overcome the effects of seed-derived PA.

Approaches as simple as following certain food processing and preparation techniques [11] to dietary supplementation with minerals [12, 13] and phytases [14-18], mutation breeding for impaired phytic acid biosynthesis [19], have proved useful to a certain extent. Golovan et al. [20] genetically modified pigs to produce heterologous phytases in their salivary glands. Crops have been genetically engineered to overexpress phytase [21, 22], reduce PA biosynthesis [23, 24], or target its transport and storage [25]. Of all these strategies, engineering crops is the most versatile and cost-effective alternative to achieve PA reduction in food grains. One of the most popular transgenic approaches is to generate plants over-expressing phytase enzyme in their seeds. However, the efficacy of this approach is limited by two main factors, first, loss of recombinant phytase activity in soil due to modification of its biochemical properties [26] and immobilization by adsorption [27], second, co-suppression induced silencing of phytase triggered by multiple gene copies [28]. Therefore, research on another major transgenic approach of targeting PA synthesis or its transport began increasing. This approach, however, demands strategic planning to produce desired outcomes. Previous studies have shown that perturbing either PA biosynthesis during early stages or its transport and storage post-synthesis both produces many downstream impacts that ultimately result in reduced plant performance, yield and nutritional quality [19]. It may thus suggest that engineering low-phytate trait by targeting PA synthesis during later stages may be prudent to avoid these negative impacts.

Tremendous efforts are currently being invested in developing genomic tools for transgenic enhancement of crops. The most recent technology of gene silencing called RNA interference (RNAi) has revolutionized customization of gene knock-downs. RNAi offers great advantages over previously discovered technologies of antisense-mediated silencing and co-suppression owing to its specific silencing, thus higher

efficiency and rapid screening [29]. It uses small double-stranded RNA molecules (dsRNA) to target the degradation of homologous mRNAs [30]. Various methods of introducing dsRNA into plant cells to successfully activate the RNAi pathway have been studied [30-35]. Tissue/organ-specific RNAi vectors have recently been proven useful to enable spatial control of gene silencing resulting in negligible interference with the plant life cycle.

Keeping in mind the above mentioned, in order to develop low PA (lpa) soybean, we propose to engineer its metabolic pathway by regulating three major factors viz. the selection of an appropriate target gene, the use of a specific promoter, and the deployment of an efficient gene silencing approach. For successfully engineering the lpa trait, firstly, we decided to select inositol polyphosphate 6-/3-/5-kinase (*GmIPK2*), one of the late PA pathway enzymes as our target gene. It is a promiscuous kinase which participates in sequential phosphorylation of inositol polyphosphates, i.e., D-*myo*-inositol 1,4,5-trisphosphate to D-*myo*-inositol 1,3,4,5,6-penta*kis*phosphate. Owing to this multiple substrate specificity, *GmIPK2* play a lead role in regulating the turnover of cellular PA levels. We, therefore, believe that if we perturb PA synthesis at this particular step, we can achieve a high reduction in PA levels. Once we confirmed *GmIPK2* as the gene target, we decided to use vector-based RNAi technology in order to achieve its efficient silencing. We considered two major factors in designing the RNAi construct, first, the type of effective small interfering RNA (siRNA) and second, the control excised on its expression. To efficiently produce siRNAs, we designed and introduced *GmIPK2* specific intron-spliced hairpin transgene in our silencing construct. Smith et al. [31] proved that inverted repeat self-complementary sequences linked by an intron are almost 100% effective in triggering RNAi. Thereafter, many research groups working with plants used hairpin RNA expression

vectors to successfully silence the corresponding target genes. Several siRNA expression vectors are commercially or publicly available. To express our transgenic hairpin RNA cassette, we chose a vector containing vicilin storage protein gene promoter with the aim to restrict PA perturbations only to phytate-accumulating seeds (since soybean cotyledon contains 90% of the seed phytate [36]). The seed-specific suppression of *GmIPK2* gene has the potential to avoid unintended effects on normal plant phenotype, and may thus result in lpa lines with a good overall performance and yield. Therefore, we based our hypothesis keeping in mind the same.

1.2 Hypothesis

The suppression of *GmIPK2* gene in the seeds of soybean cv. Pusa-16 would result in reduced PA accumulation and, a corresponding increase in available phosphorus and some PA bound minerals.

1.3 Objectives

1. Isolate *inositol polyphosphate 6-/3-/5-kinase* (*GmIPK2*) gene from cDNA of developing soybean seeds and perform its characterization by *in silico* analysis, homology modeling and molecular docking simulation.

2. Design a *GmIPK2* specific self-complementary hairpin RNA construct, capable of expressing under the control of a seed-specific promoter and perform cotyledonary node transformation of soybean with *Agrobacterium tumefacians* harbouring the RNAi construct to generate lpa transgenics.

3. Perform molecular screening of putative transgenic lines and conduct their biochemical and morphological traits analysis.

Chapter 2

Literature Review

2.1 Soybean

2.1.1 Origin, classification, and cultivation

Soybean, *Glycine max* (L.) Merr. is a member of the large botanical family of legumes, namely Leguminosae or Fabaceae. It originated in Southeast Asia, first domesticated by farmers in China from where its cultivation spread to Japan and Korea. Centuries later it was introduced to the Indian sub-continent and even later to the Americas and Europe. It is now a major crop in the United States, Brazil, Argentina, China, and India. Currently, India stands fifth in world soy production. Soybean is adaptable to varied agro-climatic conditions though its growth is optimum in a subtropical climate with mean temperatures between 20 to 30°C and soils rich in organic content such as alluvial soil. *Glycine max* is the cultivated variety of soybean, that belongs to the subgenus *Soja* of genus *Glycine* Willd, with a chromosome number 2n = 40 and an estimated size of approximately 1.1 Mbp/1C [37]. It is an annual species having several varieties with different characteristics viz. flower colour, seed colour, seed size, seed shape, seed hilum colour, plant height (cm), plant growth type, hypocotyl anthocyanin pigmentation, etc. The plant generally grows to a height of 20-180 cm, bear white or purple flowers and form clusters of pods in the leaf axils which usually contain 2-3 seeds with a yellow, green, brown, or black hull. The roots bear nodules formed by *Rhizobium japonicum* which fix nitrogen from the atmosphere into the soil in order to stimulate plant growth. Mature seeds with less than 14% moisture content is harvested typically 90-100 days after they are planted. However, the harvest

depends upon several factors such as variety, date of plantation, growing regions, local weather, etc. Generally, an active harvest is done during October and November.

2.1.2 Biochemical composition

Soybean due to its exceptional nutritional composition has immense prospects to facilitate food security and prevent malnutrition in developing (undernutrition) and developed (over-nutrition) countries.

2.1.2.1 Nutritional highlights

Protein (40-42%) and oil (18-22%) are the two major seed constituents that make soybean a significant crop. However, their percentages are influenced greatly by genetic as well as environmental factors. Its high-quality protein often referred to as a complete protein holds a Protein Digestibility-Corrected Amino Acid Score (PDCA AS) of one which is equivalent to animal protein (meat, eggs, and casein) and thus reflects its importance as a plant protein in human growth and health. Some unique peptides in soy (defensins, glycinins, conglycinins and lunasin) are known to have health-promoting effects including improved blood pressure and blood sugar regulation, better immune function, etc. The raw bean oil is rich in unsaturated fatty acids with 95% present as polyunsaturated fat, mainly linoleic acid and is a rich source of sphingolipids. The oil also contains phytosterols which can be converted into steroid hormones. Besides protein and oil, carbohydrates also account for a major 30% of dry bean weight. The majority of soy carbohydrates are insoluble complexes that constitute a good source of dietary fibre and thus make the whole beans very low on the glycemic index. The soluble carbohydrates have prebiotic properties that help promote a healthy gut microbiota. Soybeans are packed with essential minerals and

vitamins including iron, phosphorus, potassium, manganese, molybdenum, copper, zinc, magnesium, vitamin C and several B vitamins, particularly thiamine and folate. In general, 100 g of raw soybeans supply 446 kcal energy, 36 g protein, 20 g total fat, 30 g carbohydrates, and 9 g water.

2.1.2.2 Anti-Nutritional factors

Raw soybean seeds contain several biologically active factors which cause physiological damage when consumed over a long period of time. These factors with anti-nutritional (ANFs) properties include PA, proteinase inhibitors (trypsin and chymotrypsin inhibitors), alpha-amylase inhibitor, goitrogens, urease, lathyrogens, tannins, saponins, glycinins, hemagglutinin, etc. Some ANFs can be destroyed or inactivated by proper physical and chemical treatment methods or reduced to safer levels while others are unaffected by the current methods applied commercially. Amongst the humungous ANFs listed above, PA particularly summons our attention as it impairs the bioavailability of essential mineral nutrients particularly P in humans and animals who depend on soybeans and many soy-containing products and subject them to a higher risk of deficiencies of these minerals.

2.2 Phytic acid

2.2.1 Structure and function

Hartig in 1855 [38] discovered PA in tiny non-starch grains from several plant seeds. PA is a major component of mature seeds constituting around 1-5 percent of its dry weight. It functions as a reserve material for P accounting for 60-80 percent of total seed P [39, 40]. The distribution of PA in seeds and grains varies in monocot and dicot

plants. In monocots, such as rice, PA is concentrated in the germ or aleuronic layer and only traces are found in the endosperm whereas in dicot seeds such as soybean, PA is mainly found in the cotyledon. During seed development, the PA is accumulated with other seed storage compounds within globoid inclusions dispersed all through the cotyledon in case of dicots [41, 42]. Factors such as genetic variation, environmental fluctuations, and growing conditions can affect the PA content in mature grains [43]. Apart from seeds, it is found in different plant organs including roots [44], leaves [45], tubers [46], pollen, fruit etc in amount relatively low when compared with that in seeds and grains.

PA, also known as *myo*-inositol hexa*kis*phosphate (InsP6), is the hexaphosphoric ester of hexahydric cyclic alcohol *myo*-inositol (Figure 2.1). For many years during the twentieth century, its structure has been intensively discussed in the chemical society [47]. In mature plant seeds, PA due to its anionic nature exists in the form of complex mineral salts known as phytin [42] and in some cases, bind to proteins and starches.

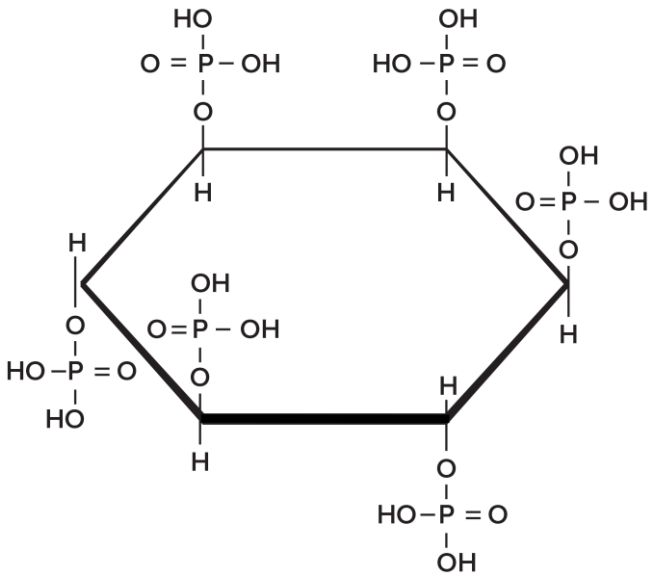


Figure 2.1: Chemical structure of phytic acid.

In seeds, phytin serves: 1) as a reservoir of P, 2) as a storehouse of energy, 3) as an adversary for adenosine triphosphate production through phytin biosynthesis at seed maturity, 4) as a chelator of bivalent mineral cations required for regulating cellular processes, and 5) to modulate the level of inorganic phosphorus (Pi) in the seeds [48]. It supports seedling growth during germination by undergoing phytase degradation to fulfil the above-stated functions. Research over the years has shown that PA pathway enzymes and metabolites play regulatory roles in cell signalling [49, 50] and plant processes including membrane trafficking [51], stress response [52-54], P homeostasis [55], photomorphogenesis [56], and mRNA nuclear export [57].

Besides plants, PA is a ubiquitous cellular component in almost all the mammals where it performs numerous regulatory functions [58]. It plays an important role in regulating cellular apoptosis [59, 60], imparts neuroprotection [61], and function as an enzyme cofactor [62]. Investigative research in yeast have suggested that it also play an essential part in nuclear export of mRNA [63], DNA repair [64], and regulation of chromatin structure [65, 66]. PA has strong antioxidant properties with the potential to alleviate the risk of certain types of cancers [67]. It also finds applications in reducing the risk of cardiovascular diseases, osteoporosis, diabetes and reducing menopausal symptoms [68]. For these reasons, speculating the role of dietary phytate in human health and nutrition is of foremost concern.

2.2.2 Anti-nutritional properties

There is growing interest in seed PA due to its significant role in influencing human health and nutrition. Studies have shown that phytate in humans and animals affects nutrient bioavailability. The deplorable effect can be accounted to its chemical structure which provides 12 reactive sites (replaceable protons). Six of these sites

are strongly acidic in nature [69] and carry a negative charge at the physiological pH which makes it capable of binding polyvalent mineral cations such as Zn^{2+}, Fe^{2+}/Fe^{3+}, Ca^{2+}, Mg^{2+} into stable complexes (Figure 2.2) [70-75].

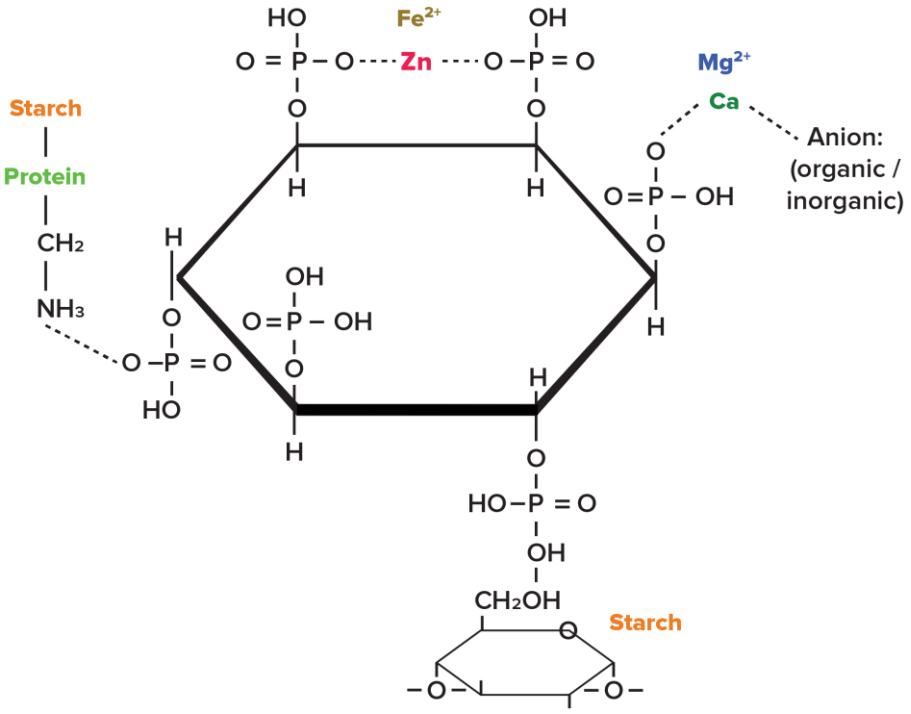


Figure 2.2: Interaction of phytic acid with mineral cations, protein, and starch.

These complexes are highly insoluble in nature and are readily hydrolyzed by phytases [5]. Since monogastrics lack phytases in their digestive tract, they show poor PA digestibility [70, 76]. This consequently diminishes the nutritive value of the seeds by restricting the release and hence bioavailability of P and other bound minerals. For this reason, mineral deficiencies are often observed in populations which rely on cereal and legume as staple grains [77]. Iron and zinc contribute to the most widespread PA associated micronutrient deficiencies often referred to as 'hidden hunger' (UNICEF, 1990). Amongst the number of cases of rickets reported in people belonging to Pakistan, Northern India, Iran, etc., a strong correlation was

discovered with their diets rich in PA particularly in the form of chapatis [78-80].

In addition to chelating important minerals, PA also binds positively charged terminal amino groups of proteins like lysyl, histidyl and also inhibits essential digestive enzymes pepsin, carboxypeptidase, trypsin, amylase [81]. The binding of PA with starch has also been reported which may reduce its availability from the diet [48].

2.2.3 Environmental impact

In areas of intense livestock production in order to fulfil animal's nutritional requirements, the feeds are supplemented with inorganic P or commercial *phytases*. Manure derived from these animals is subsequently applied as fertilizer to the croplands. A wide range of *phytases* inherent to the soil microorganisms contributes to hydrolyze the phytates present in livestock litter [82]. This results in a continuous P supply to the fields in amounts greater than that can be deployed in the agricultural produce and create an imbalance in the soil P levels causing it to rise. Lott and co-workers [83] estimated that PA phosphorus from crops amounts to nearly 65% of the world annual P fertilizer share. This eventually results in P pollution of agricultural systems and stimulates or accelerates eutrophication of water bodies, raising a serious environmental concern [84].

2.3 Phytic acid biosynthesis pathway

PA biosynthesis have been studied in a number of organisms including mungbean [85], duckweed [86], *Arabidopsis* [87, 88], rice [89], soybean [90], common bean [91], wheat [92], slime mold [93], yeast [94], *Drosophila* [95], and human [96]. Two alternative routes have been suggested for its formation: a "lipid-dependent" pathway,

and a "lipid-independent" pathway (Figure 2.3) [86, 97-100].

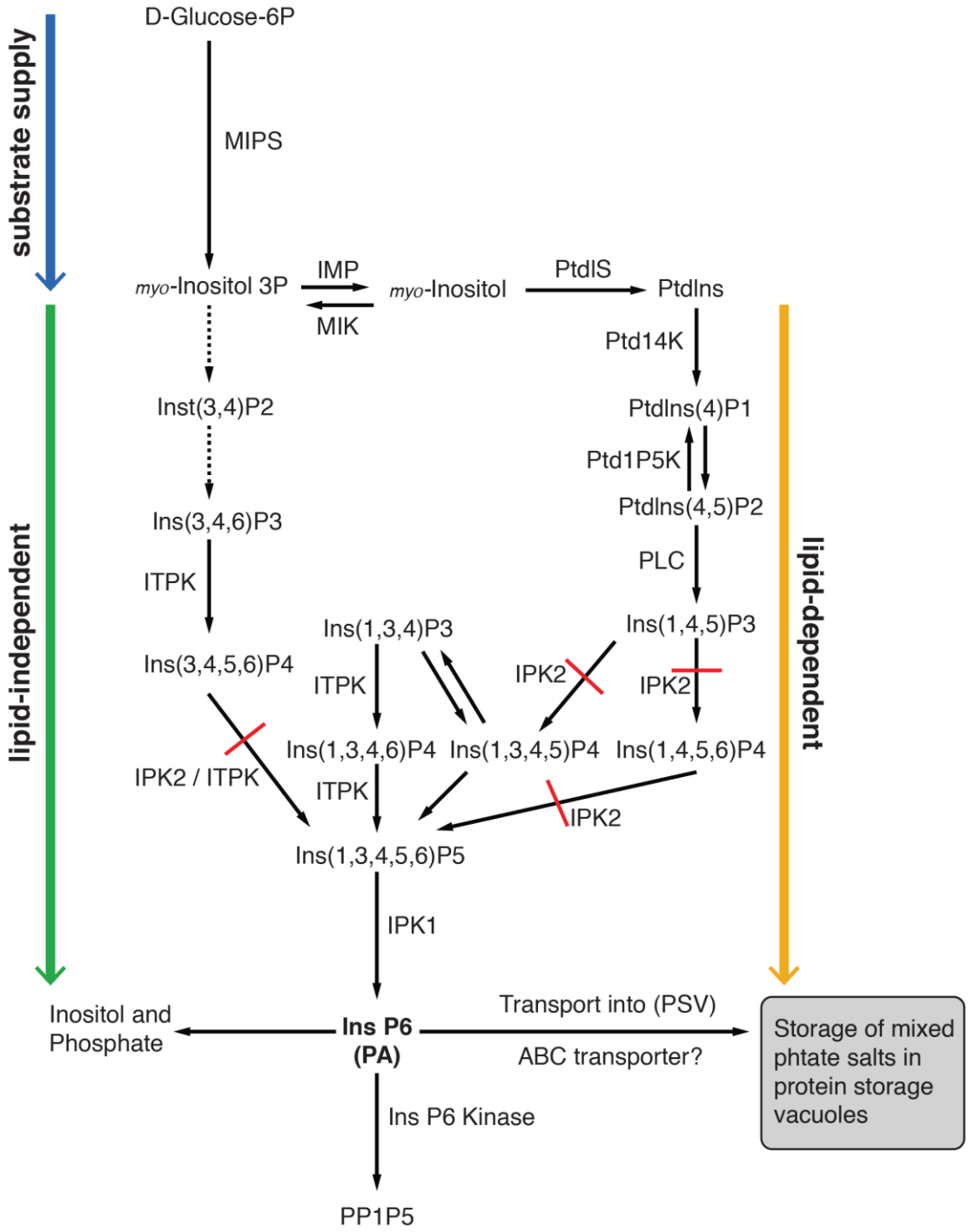


Figure 2.3: A schematic diagram of phytic acid biosynthesis pathway in plants via both lipid-independent (green) and lipid-dependent (orange) routes. Substrate and enzyme abbreviations are elaborated under section 2.3. Red line indicates the pathway steps distrupted in our study to generate lpa soybean.

In plants, the former pathway has been characterized in almost all the tissues while the later appears to predominate in seeds. In this section, we provide a brief review of the important genes and metabolites involved in PA synthesis.

2.3.1 Early pathway

The first step in PA biosynthesis pathway via both the routes stated above is the same i.e. the de novo production of *myo*-inositol. The D-*myo*-inositol 3-phosphate synthase (MIPS) enzyme catalyzes the synthesis of *myo*-inositol by converting D-glucose 6-phosphate to *myo*-inositol 3-phosphate (Ins(3)P1). The number of loci encoding *MIPS* gene varies across species [86, 91, 100-104]. Soybean contains four *MIPS* isoforms of which the *GmMIPS1* isoform is expressed most strongly in seeds [21]. Ins(3)P1 is then either phosphorylated to *myo*-inositol 3,4-phosphate (Ins(3,4)P2) by the action of phosphoglycerate kinase (PGK) in lipid-independent pathway or hydrolyzed to free *myo*-inositol and P by enzyme inositol monophosphatase (IMP) in lipid-dependent pathway in a reversible manner to re-supplement lipid-independent pathway by the action of enzyme *myo*-inositol kinase (MIK). In lipid-dependent pathway, free *myo*-inositol so generated is sequentially processed to phosphatidylinositol (PtdIns), phosphatidylinositol 4-phosphate (PtdIns(4)P1), and phosphatidylinositol 4,5-bisphos-phate (PtdIns(4,5)P2) by phosphatidylinositol synthase (PtdIS), phosphatidylinositol 4-kinase (PtdI4K) and phosphatidylinositol 4,5-bisphosphate kinase (PtdIP5K) respectively. PtdIns(4,5)P2 is finally hydrolyzed by phospholipase C to yield Inositol 1,4,5-trisphosphate (Ins(1,4,5)P3).

2.3.2 Late pathway

Inositol triphosphates (InsP3s) generated from both the pathways are sequentially phosphorylated to PA by three main classes of inositol kinases, namely inositol

polyphosphate 6-/3-/5-kinase (IPK2), inositol 1,3,4-trisphosphate 5-/6-kinase (ITPK), and inositol polyphosphate 2-kinase (IPK1), which together constitute this part of the pathway. IPK2 and ITPK catalyze the conversion of inositol triphosphates to inositol penta*kis*phosphate. IPK2 kinase is known to participate specifically in the lipid-dependent pathway which is not suspected to be the major route for PA synthesis in the seeds. Previous studies although have shown that its activity is important to seed PA synthesis [87]. During seed development, its transcripts show the highest accumulation after that of *MIPS* and *IMP* [105]. IPK1 catalyzes the final conversion of inositol penta*kis*phosphate to PA. Like *MIPS*, the number of loci encoding this gene also varies across species. Amongst three of its orthologs identified in soybean, only one shows high expression in the seeds [106]. PA once synthesized in the cytoplasm is actively transported by a multidrug-resistance-associated ATP-binding cassette (ABC) transporter protein and ultimately stored as mixed salts inside the protein storage vacuoles. Research has shown that it plays a significant role in regulating PA levels in the seeds [25, 107].

2.4 *IPK2*: Late pathway *myo*-inositol kinase

The *inositol polyphosphate 6-/3-/5-kinase* gene encodes a multifunctional enzyme that carries two-step phosphorylation of *myo*-inositol trisphosphate (InsP3) to *myo*-inositol penta*kis*phosphate (InsP5) via a *myo*-inositol tetra*kis*phosphate (InsP4) intermediate. Its 6/3/5-kinase activity was reported by Stevenson-Paulik and co-workers [108]. They identified two isoforms of the gene, *AtIpk2α* and *AtIpkβ* in *Arabidopsis* that were able to phosphorylate InsP3 with a Km of 14.6 μM and InsP4 with a Km 15.5 μM and 31.7 μM. In yeast, *IPK2* reportedly play an important role in PA biosynthesis and its disruption has shown to lower the production of PA by a

100-fold [109]. Besides its kinase activity, IPK2 in yeast is also an integral part of ArgR-Mcm1 transcriptional complex which controls arginine metabolism [94]. Apart from *Arabidopsis* and yeast, *IPK2* has been characterized in other organisms including human [110, 111], rat [112, 113], and chicken [114].

2.5 Strategies for regulating dietary PA consumption

2.5.1 Food preparation and processing techniques

Several efforts are underway to overcome the negative health effects (highlighted above) associated with plant-derived dietary PA consumption. Detailed studies have been done in the past on food processing methods used to reduce the PA content of plant-based foods. Processes involving enzymatic hydrolysis of PA such as germination, malting, and fermentation [115-118] and non-enzymatic methods such as soaking, milling, thermal processing [119-122] have shown to reduce the concentration of PA and improve other minerals bioavailability in some plant staples. These strategies however have some disadvantages such as significant loss of vitamins & minerals found in the aleurone layer or germ in case of milling, loss of water-soluble minerals and proteins during soaking and germination, accumulation of lower inositol phosphates during germination and fermentation, altered chemical nature of the feed in case of thermal processing, etc.

2.5.2 Dietary supplementation

2.5.2.1 Mineral supplementation

One of the most common ways to make up for the loss of PA associated minerals is through dietary supplementation. Sodium iron ethylenediaminetetraacetic acid [NaFe

(III)EDTA], has been successfully used as a dietary fortificant in developing countries [12, 123-125]. Animal feeds are also supplemented with an available form of P and other minerals [13]. However, such food fortification programs are widely practiced on industrially processed food items. Their success, therefore, relies on the availability of infrastructure, purchasing power, access to markets and healthcare systems, which can be afforded by only fifty percent of the world's population [126].

2.5.2.2 Phytase supplementation

Dietary supplementation with exogenous phytase is another widely used strategy to ameliorate bioavailability of P and other minerals [127]. Phytases are a special class of phosphatases which sequentially hydrolyze phosphate ester bonds of PA [128] to yield inorganic phosphorus (Pi) and *myo*-inositol or a succession of lower esters of phosphoric acid [129]. Microbial phytases are frequently used enzyme supplements in the diets of monogastric animals [130]. The first commercial phytase was isolated in 1991 by Gist-Brocades from a genetically modified strain of *Aspergillus niger* and marketed under the brand name Natuphos. Since then, other phytase preparations such as RonozymeP, Phyzyme XP, OptiPhos, Finase, Quantum, etc. were successfully launched in the market. Li and co-workers [131] expressed *A. niger* gene *PhyA* from suspension cultures of soybean. Denbow and co-workers [132] evaluated the effectiveness of supplemental phytases by conducting an experiment with male broilers fed on supplemented soybean meal and found that 31 to 58% of phytate P was released and made available. In addition to a reduction in supplemental Pi, exogenous phytase results in a concomitant reduction of fecal P excretion [133]. Vats and Banerjee, 2004 from their study suggested replacement of Pi supplementation equivalent to 1 g with 500-1000 units of phytase enzyme to improve bioavailability

and simultaneously reduce P excretion by 30-50% [134]. Phytase in feeds also increases the feed conversion ratio for the animals thus improving their growth performance [135]. However, the use of phytases in food and feed industries call for some critical challenges. When part of a formulation, they can be inactivated by several factors in its surroundings such as temperature, pH, etc. [136-138]. Apart from this, phytases are recognized as occupational allergens with the potential to cause immune responses specific to IgE.

2.5.3 Developing low phytic acid crops

2.5.3.1 Mutation breeding

Another promising strategy is to generate low phytic acid (lpa) phenotype [139] by inducing mutagenesis. Several lpa mutations induced by radiations and chemicals have been segregated by phenotypic screening in major grain crops (Table 1). Downregulation of PA synthesis, transport, and storage, genes have been traced to most of these mutants. Genotypic characterization traced lpa mutations into three different categories depending upon the PA biosynthesis gene it targets: 1) early pathway mutations (*MIPS, MIK, IMP*) mainly altering the first step (*MIPS*, D-glucose 6-phosphate to *myo*-inositol 3-phosphate) of PA synthesis thereby affecting substrate supply, 2) mutations perturbing late pathway genes (*ITPK, IPK2, IPK1*), 3) mutations affecting PA transport and storage (*MRP*). Lpa mutations are normally distinguished by reduction in PA levels and a simultaneous equivalent increase in Pi. Unfortunately, a majority of lpa mutants report adverse agronomic traits, such as poor seed germination and in extreme cases seed abortion, stunted plant growth, low seed weight, etc. which make them unappealing to the plant-breeders [148, 157-160].

Table 1: Low phytic acid mutations in legume and cereal crops.

Species	Locus	Target gene	Seed PA reduction (%)	Metabolites regulated	References
Soybean	*pha1 & pha2*	MRP	80	Concomitant increase in Pi	Wilcox et al., 2000[140]; Oltmans et al., 2004 [141]
	lpa-ZC-2	IPK1	50	Concomitant increase in Pi, molar increase of lower InsPs	Yuan et al., 2007 [142]
	LR33	MIPS	50	60-80% decrease in *myo*-inositol	Hitz et al., 2002 [100]
	lpa-TW75-1	MIPS	50	Decrease in *myo*-inositol	Frank et al., 2009 [143]; Yuan et al., 2007 [142]
Barley	*lpa1*	IMP	50	Endosperm/aleurone specific, increase in Pi, ~15% reduction in total seed P	Larson et al., 1998 [144]; Ockenden et al., 2004 [145]
	lpa2	IMP	50	Concomitant increase in Pi, increased lower InsPs and *myo*-inositol	Dorsch et al., 2003 [146]; Roslinsky et al., 2007 [147]
	lpa3	IMP, MIPS	70	Concomitant increase in Pi, increased *myo*-inositol	Larson et al., 1998 [144]; Roslinsky et al., 2007 [147]
Maize	*lpa1*	MRP	50 to > 90	Allele-specific, concomitant increase in Pi, increased *myo*-inositol	Raboy et al., 2000 [148]; Pilu et al., 2003 [149]
	lpa2	ITPK	50	Concomitant increase in Pi, increased lower InsPs and *myo*-inositol	Shi et al., 2003 [43]; Pilu et al., 2003 [149]
	lpa3	MIK	50	Largely embryo specific, concomitant increase in Pi, increased *myo*-inositol	Raboy et al., 2000 [148]
Rice	*lpa1*	2-PGK	40	Concomitant increase in Pi, molar increase of lower InsPs	Larson et al., 2000 [150]; Liu et al., 2007 151]
	lpaN15-186	MIK	75	Concomitant increase in Pi, increased *myo*-inositol	Kim et al., 2008 [152]; Xu et al., 2009 [107]
	lpa-XS110-1	MIK	64	Concomitant increase in Pi, increased *myo*-inositol	Frank et al., 2007 [153]; Zhao et al., 2008 [154]
	lpa-XS110-2	MIK	20	Decreased *myo*-inositol	Xu et al., 2009 [107]; Liu et al., 2007 [151]

Table 1: Low phytic acid mutations in legume and cereal crops.

Species	Locus	Target gene	Seed PA reduction (%)	Metabolites regulated	References
Rice	*lpa-XS110-3*	MIK	100	Increased *myo*-inositol, lethal phenol-typic alterations	Liu et al., 2007 [151]; Xu et al., 2009 [107]
	XQZ-1	2-PGK	40	Concomitant increase in Pi, molar increase of lower InsPs	Zhao et al., 2008 [154]
Common Bean	*lpa1*	MRP	90	MIPS, IMP, IPK2, ITPKα, IPK1 genes downregulated	Panzeri et al., 2011 [155]
Garden Pea	*1-150-81; 1-2347-144*	unknown	60 to 65	Molar increase in Pi	Warkentin et al., 2012 [156]

These negative effects are expected since many lpa mutations primarily disturb cellular inositol metabolism that is essential to support normal plant progression. The strength of pleiotropic effects observed in lpa mutants is a result of the number of lpa loci affected. In addition, a number of studies described metabolic crosstalk between PA and signalling pathways which may result in unexpected phenotypes in mutants that result in PA reductions of more than 70% [161-164]. Researchers are therefore seeking a different approach to overcome the severe pleiotropic problems associated with lpa mutation breeding.

2.5.3.2 Genetic engineering

Genetic manipulation of the PA biosynthetic pathway is being explored, wherein different enzymes involved in the pathway are regulated to reduce phytate content. This approach of generating lpa crops provide several advantages over random "whole-plant" mutations discussed in the above section.

2.5.3.2.1 Expressing recombinant phytase in plant seeds

Genetic engineering has enabled researchers to introduce a stable phytase-encoding gene in the desirable seed tissue, which is active during both seed development as well as in matured seeds. Studies have proven that microbial phytases being most stable are the most promising phytases for such biotechnological applications [165]. Transgenic seeds containing a reserve of these phytases provide a high phosphate and mineral nutritional value than conventional phytase supplemented feeds and reduce P excretion by providing for better utilization of phytin reserves stored in the seed. These recombinant microbial phytases have been expressed in several plants. Pen and co-workers [166] demonstrated that seeds of transgenic tobacco plant showing

constitutive expression of recombinant *phytase PhyA* of *Aspergillus niger* fused with signal peptide PR-S protein is a source of apoplastic phytases capable of releasing free P from dietary phytate. Later, Denbow and co-workers [16] also constitutively engineered recombinant *A. niger PhyA* in soybean under CaMV 35S promoter. Feeding trials were subsequently conducted by combining standard diets with recombinant tobacco and soybean seeds containing engineered *phytases,* in order to digest dietary phytate, and improve its nutritional accessibility and simultaneously reduce the level of Pi excreted. Transgenic alfalfa and canola seeds expressing apoplastic *PhyA* have been patented as an important feed supplement [22]. Drakakaki and co-workers [167] and Chen and co-workers [168] successfully expressed *A. niger PhyA* driven by an endosperm-specific promoter in maize. Rice [169] and Wheat [170] expressing seed fungal phytases have also been generated. Besides fungal, phytases from other microbial sources have also been expressed in transgenic crops. Hong and co-workers [171] expressed bacterial phytases in the germinated rice seeds while Hamada and co-workers [172] expressed yeast phytase in its leaves. Bilyeu and co-workers [22] transformed *Escherichia coli appA* gene in soybean and observed that the PA levels were reduced in the mature seeds by more than 90%. Apart from microbial phytases, phytases from plants are also utilized. Chiera and co-workers [173] cloned *GmPhy* [21] in soybean under promoter β-conglycinin specific for seeds and reduced the phytate content in its seeds by 8%. Besides the stated advantages, this approach faces a major drawback. The seeds from these transgenic crops before feeding to the animals must first be processed by grinding and incubation in water at 37°C in order to activate the enzyme. This thus adds to high processing and labor cost. Therefore, in order to circumvent this approach, a simple, sustainable, and a cost-effective solution is required.

2.5.3.2.2 Silencing endogenous phytic acid biosynthesis genes

Gene silencing is an essential mechanism to regulate gene expression in eukaryotes. In silencing, several small RNA molecules lead sequence-specific inhibition of either gene transcription or translation [174-177]. RNA silencing plays a significant role in regulating chromatin structure, genome stability, defense against viral infections and other foreign genetic material and other important biological processes [178-181]. While it is an evolving class of mechanisms, the most common and well-studied amongst them is RNA interference (RNAi). It is the most powerful and efficient dominant gene silencing approach [31] at times resulting in complete silencing of the gene as observed in a knockout approach. It was initially employed as a tool to manipulate metabolic pathways by targeting the suppression of their genes. It instantly gained remarkable attention owing to its successful application in genomic-scale studies. It holds great potential in the field of clinical research and therapeutics [182], to design new drugs against different disease targets in animal and humans [183, 184]. In plant sciences, RNAi technology has been successfully employed for genetic improvement of crops [185-187] such as to enhance their nutritional value, to confer resistance against diseases and certain pathogens, etc. [184]. RNAi manipul-ation of PA biosynthetic pathway has been explored in many crops and can, therefore, be explored in soybean crop whereby different enzymes involved in the pathway can be down-regulated to reduce seed phytate content.

2.6 RNA interference

2.6.1 Discovery

In the early 1990s, scientists in the United States made accidental discovery of RNAi in unanticipated outcomes of experiments performed in plants. In 1990, Napoli and

co-workers [188] were working on enhancing the pigment coloration of commercial petunia flowers by constitutively over-expressing the *CHS* gene. To their surprise, the flowers instead of having a deep purple color unexpectedly appeared variegated and even white. They observed that the introduction of *CHS* gene silenced both the endogenous and transgenic genes, resulting in a phenomenon termed co-suppression. Romano and Macino [189] and Cogoni and co-workers [190] reported a similar phenomenon named quelling in the fungus *Neurospora crassa*. Andrew Fire and Craig Mello later discovered small RNA molecules in *Caenorhabditis elegans* which evidently suppressed the target gene expression and caused their silencing [191-193]. They named this kind of gene silencing as "RNA interference" (RNAi). Guo and Kemphues [194] later discovered by their investigations in *C. elegans* that a gene can be successfully silenced by injecting either its corresponding sense or antisense RNA molecule. Waterhouse and co-workers [195] found that gene silencing can be induced more efficiently by expressing sense and antisense RNAs concurrently. Considering the findings made for the gene silencing alternatives discussed above it was later revealed that they all lead to the activation of RNAi by a common route of generating double-stranded RNAs (dsRNAs) to silence homologous loci [196-199].

2.6.2 Basic mechanism

The general process of RNAi is triggered by endogenous or exogenous dsRNA molecules (viral genome, transgene or other foreign genetic element) sharing exclusive sequence homology to the target gene, delivered in the cell [200-202]. dsRNAs can form by sense and antisense hybridization, folding back of inverted-repeat sequences, hybridization of foreign RNA molecules showing sequence complementarity or synthesis by RNA-dependent RNA polymerases (RDRs). Two

functional types of RDRs are recognized in plants RDRα and RDRγ [203]. Following the initial identification of dsRNAs, the overall process of RNAi (Figure 2.4) can be divided into three stages:

1) Generation of small RNAs

Long dsRNAs in the cell that are either endogenously produced or delivered by scientific experiment are converted to small RNAs by a ribonuclease III enzyme, Dicer. Dicer polypeptide share four key domains: a dsRNA binding domain, a helicase domain specific to the amino terminal, a PAZ (Piwi/Argonaute/Zwille) domain and RNase III domain [204-206]. It binds the 2-nucleotide 3′ overhang of dsRNAs through its PAZ domain and cleaves them to 21-25 bp double-stranded small RNAs (sRNAs) [178, 207-209]. sRNAs in eukaryotes can be divided into several major classes including micro RNAs (miRNAs), small interfering RNAs (siRNAs), piwi-interacting RNAs (piRNAs), small temporal RNAs (stRNAs), etc. Amongst them, siRNAs and miRNAs have been well characterized in plants [210-211]. Several classes of siRNAs have been recognized including hairpin-derived siRNAs (hpRNAs), natural antisense siRNAs (natRNAs), secondary siRNAs, heterochromatic siRNAs (hetRNAs), trans-acting siRNAs (TaRNAs) and repeat-associated siRNAs (RaRNAs), based on their biogenesis and functions. All sRNAs in plants are 2′-O-methylated at the 3′ end to confer stability by protection from 3′ uridylation and degradation [212]. We will discuss siRNAs induced RNAi mechanism in the remaining section as it is the chosen mode for our study.

2) Silencing complex formation

Following production, Dicer incorporates siRNAs into a multiprotein endonuclease complex called the RNA-induced silencing complex (RISC). The Argonaute protein,

AGO forms an integral part of this complex. AGOs have characteristic sRNA binding PAZ domain which recognizes its 3' end and a PIWI domain essential to induce cleavage of target mRNA [213, 2141]. In RISC, AGO combines with siRNAs to form its core. X-ray crystallographic analysis revealed that the phosphorylated 5' end of siRNAs forms a nucleation site by divalent interactions (such as magnesium) and aromatic stacking of the nucleotide with a conserved active site tyrosine residue [215, 216]. The AGO associated siRNAs serve as guides for sequence-specific recognition of target messenger RNAs. In other words, these AGO proteins direct mature siRNAs to its target mRNA.

3) Activation of silencing complex and mRNA degradation

Finally, siRNA duplex unwinding is performed by the helicase domain of RISC in an ATP-dependent process [217]. RISC then distinguishes between the two separated siRNA strands and retains only the antisense strand also known as the guide strand, degrading the passenger (sense) strand [218]. The guide strand directs RISC (with nuclease activity) to homologous mRNAs and induces cleavage of their phosphodiester bonds resulting in mRNA degradation [204, 215, 219], or causes suppression of their translation, collectively called post-transcriptional gene silencing (PTGS). Following complete cleavage, the RISC and siRNA complex is free to participate in a fresh round of target mRNA recognition and cleavage [220]. The RNAi effect is further amplified by RDRs which supplement the existing pool of siRNAs by using them as primers to synthesize new siRNAs [221, 222]. Besides PTGS, siRNAs also guide transcriptional gene silencing (TGS) by participating in DNA methylation and histone modification in the nucleus [181].

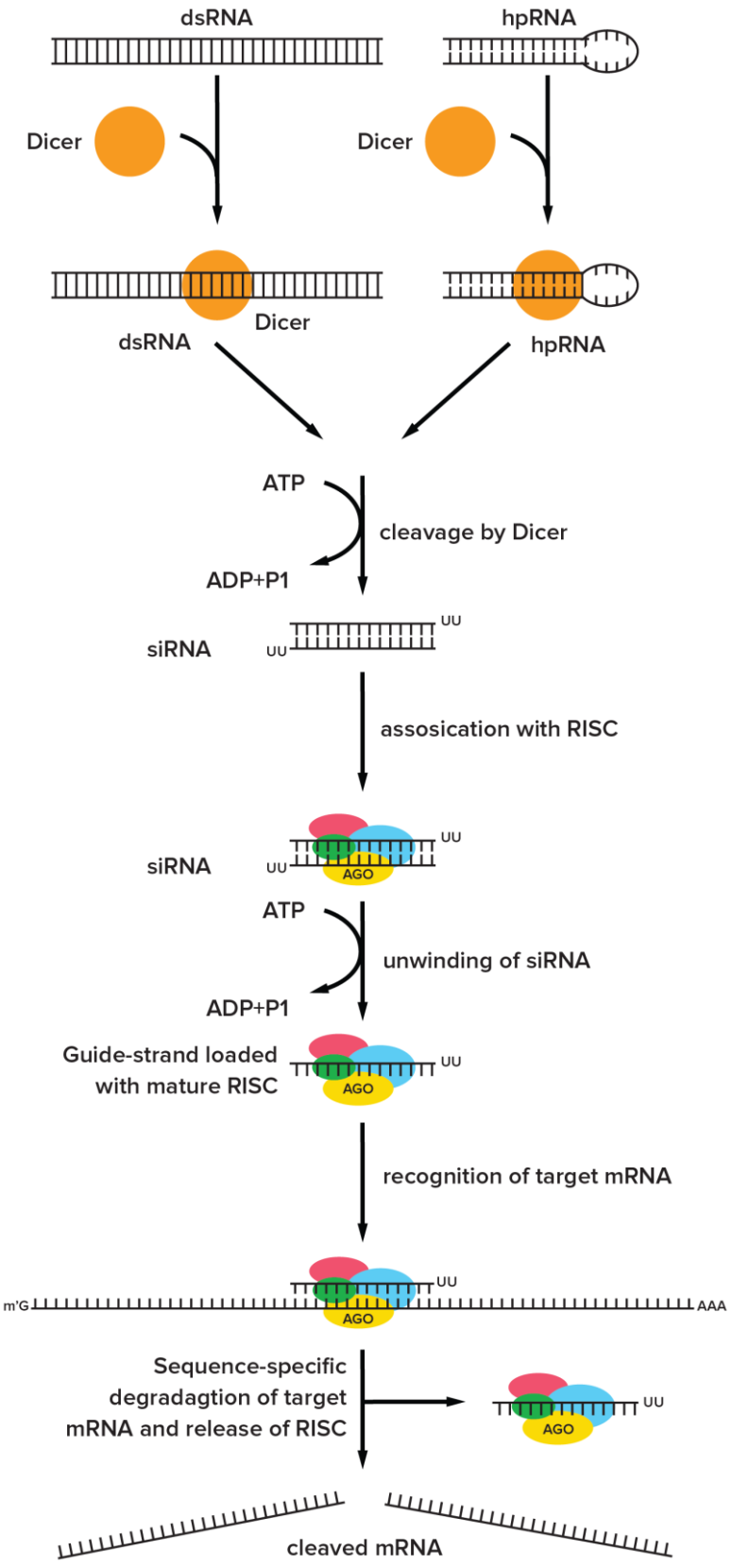


Figure 2.4: Diagram showing the basic mechanism of RNA interference pathway.

2.6.3 Applications in crop improvement

RNAi technology is drawing growing attention of plant molecular biologists throughout the world due to its direct or indirect effect on several plant physiological and biochemical processes which collectively control many important agronomic traits. It has been effectively exploited for altering plant architecture, developing biotic and abiotic stress resistance, in development of male sterile plants, engineering secondary metabolites, nutritional improvement, prolongation of fruit shelf-life, seedless fruit development, removal of toxic compounds, etc. [223].

Recently, RNAi based techniques are attracting focus to improve the nutritional value of many significant plant species by manipulating their metabolic pathways [224]. Common nutritional targets include micronutrients, vitamins, fatty acids, amino acids, and antioxidants. In cotton (Gossypium spp.), hpRNA regulated silencing was successfully employed to downregulate D9-desaturase and x6-desaturase encoding genes and consequently achieve elevation in oleic (77%) and stearic (40%) fatty acids content for enhancing human health [225]. In potato, β-carotene hydroxylase gene (BCH) was silenced to successfully enhance its β-carotene and lutein content and serve as an excellent source of vitamin A [226]. Davuluri and co-workers [227] improved the carotenoid and flavonoid levels in tomatoes by expressing the RNAi construct targeting DE-ETHIOLATED1 (DET1) gene under a fruit-specific promoter. Similarly, the rapeseed (Brassica napus) carotenoid (3-carotenes: zeaxanthin, violaxanthin, and lutein) content was also improved by downregulating lycopene epsilon cyclase (e-CYC) gene expression [228]. The lysine content of major cereals and legumes can be improved with the help of RNAi technology. A lysine-rich maize variety was developed by Houmard and co-workers [229] by suppressing lysine-ketoglutarate reductase (LKR) and saccharopine dehydrogenase (SDH) gene in its

endosperm.

Further, Regina and co-workers [230] reported an enhanced synthesis of amylose (>70%) in wheat by RNAi-mediated suppression of SBE isozymes under a seed-specific promoter. A similar increase in amylose content was reported in sweet potato [231] and barley [232] as well. Ogita and co-workers reduced the caffeine content of coffee by up to 70% by RNAi induced suppression of theobromine synthase gene [233-234]. In 2005, Hüsken and co-workers used RNAi to inhibit UDP-Glc:sinapate glucosyltransferase gene responsible for bitterness in *Brassica napus* [235]. Sunil Kumar and co-workers reduced gossypol content in cotton seeds by expressing δ-cadinene synthase gene RNAi construct under a promoter specific for seeds [236]. These seeds can be used as a source of protein and calories in countries suffering from malnutrition and starvation. A company named J.R. Simplot developed Innate potato by using RNAi to silence four different proteins in five popular potato varieties including Ranger Russet, Russet Burbank, and Atlantic potatoes to resist blackspot, bruising and browning. The company also claims that Innate potato produces up to 70% less acrylamide, a potential human carcinogen than other potatoes cooked at the same temperature [237]. Jorgensen and co-workers achieved a 92% reduction of cyanogenic glucosides in cassava tubers by RNAi induced suppression of cytochrome P450 genes (CYP79D1 and CYP79D2) responsible for producing linamarin and lotaustralin [238].

2.7 RNA silencing in soybean

Soybean is an exceptionally difficult crop species to investigate gene function and regulation because it has a large, complex genome with number of gene families and highly duplicated regions [239, 240]. However, efficient RNAi cloning vectors along

with recent advancements in soybean transformation technology [241], and soybean expressed sequence tag (EST) databases & gene chips have augmented global gene function and expression analyses and enabled genetic improvement of soybean. Novel traits have been engineered in soybean through gene expression or RNAi-mediated gene silencing using transgenes as well as viral vectors [242]. Different metabolic pathways have been manipulated to accumulate valuable nutritional metabolites in its seeds and/or divert the undesirable ones (Table 2). Nunes and co-workers [23] and Shi and co-workers [25] reduced PA content up to 95% by silencing *MIPS1* and *ABC* transporter respectively and improved P availability in soybean seeds. Takagi and co-workers reduced saponin biosynthesis by RNAi targeted silencing of *β-amyrin* synthase gene [252]. Herman and co-workers achieved reduction in the accumulation of a major allergen in soybean seeds by utilizing transgene-induced RNAi without any phenotypic alterations [253]. Soybean seed fatty acid composition has also been modified to lower its polyunsaturated fatty acid (PUFA) content for stable seed oil. Wang and Xu designed a silencing construct carrying an inverted-repeat fragment of the fatty acid desaturase (*GmFAD2*) gene and obtained high oleic acid content (71.5-81.9%) [245]. Flores and co-workers effectively carried out siRNA-mediated silencing of the *FAD3* gene family in soybean in order to regulate the synthesis of ω-3 polyunsaturated α-linolenic acid [244]. Yu and co-workers engineered soybean to increase isoflavone content in its seeds [247].

2.8 Strategy for RNA interference of *IPK2* gene in soybean seeds

Earlier results of RNAi were obtained by, the transformation of plants with antisense and sense (co-suppression) RNA producing constructs. However, Chuang and Meyer-owitz observed that RNAi could be more efficiently triggered using a transgene

designed to express an inverted repeat (IR) sequence of the desired gene [262].

Table 2: Metabolic engineering via RNAi silencing in soybean.

Target gene	Promoter	Transformation method	Tissues assayed	Silencing effect	References
Myo-inositol-1-phosphate synthase, *MIPS*	CaMV 35S	Biolistics	Seed	Reduced phytic acid and in some lines seed abortion	Nunes et al. 2006 [23]
ATP-binding cassette, *ABC*	Kti3	Biolistics	Seed	Reduced phytic acid	Shi et al. 2007 [25]
Fatty acid desaturase, *FAD2-1*	β-conglycinin	Biolistics	Seed	Increase in oleic acid content	Kinney 1996 [243]
Fatty acid desaturase, *FAD3*	Glycinin	*A. tumefaciens*	Seed	Reduced linolenic acids	Flores et al. 2008 [244]
Fatty acid desaturase, *FAD2*	Lectin	*A. tumefaciens*	Seed	Increase in oleic acid content	Wang and Xu 2008 [245]
Fatty acid desaturase, *FAD2-1* Intron	β-conglycinin α′ subunit	*A. tumefaciens*	Seed	Changes in fatty acid composition	Wagner et al. 2011 [246]
Flavanone 3-hydroxylase	Kti3	Biolistics	Seed	Increased isoflavone	Yu et al. 2003 [247]
Chalcone synthase, *CHS6* and isoflavone synthase, *IFS2*	CsVMV	*A. rhizogenes*	Hairy root	Reduced isoflavone and coumesterol	Lozovaya et al. 2007 [248]
MYB transcription factor gene *GmMYB176*	CaMV 35S	*A. rhizogenes*	Hairy root	Reduced isoflavonoids	Yi et al. 2010 [249]
β-conglycinin α and α′ subunit	β-conglycinin	Biolistics	Seed	Changed protein composition	Kinney et al. 2001 [250]
Glycinin, *A1bB2* and *FAD-2*	Glycinin	Biolistics	Seed	Changed protein composition	Schmidt et al. 2011 [251]
β-amyrin synthase, *GmBAS1* and *GmBAS2*	β-conglycinin	Biolistics	Seed	Reduced saponin	Takagi et al. 2011 [252]
Gly m Bd 30 K	β-conglycinin α subunit	Biolistics	Seed	Reduced Gly m Bd 30 K	Herman et al. 2003 [253]
Thioredoxin	CaMV 35S	*A. rhizogenes*	Hairy root	Suppression of root nodule development	Lee et al. 2005 [254]

Table 2: Metabolic engineering via RNAi silencing in soybean.

Target gene	Promoter	Transformation method	Tissues assayed	Silencing effect	References
Ecto-apyrase, GS52	FMV	*A. rhizogenes*	Hairy root	Suppression of root nodule development	Govindarajulu et al. 2009 [255]
FW2.2-like gene GmFWL1	FMV	*A. rhizogenes*	Hairy root	Suppression of root nodule development	Libault et al. 2010 [256]
Delta 15 desaturase	β-conglycinin α′ subunit	Biolistics	Seed	Production of arachidonic acid	Chen et al. 2006 [257]
Seed oil body protein 24-kDa oleosin	Oleosin 24-kD isoform A	Biolistics	Seed	Changes in seed oil body size and slow growth of the plant	Schmidt and Herman 2008 [258]
Glutathione S-transferase gene GST9	CsVMV	*A. rhizogenes*	Hairy root	Reduced nitrogenase activity and increaseed oxidation damaged proteins	Dalton et al. 2009 [259]
Amino aldehyde dehydro-genase	CaMV 35S	*A. tumefaciens*	Callus	Biosynthesis of 2-acytyl-1-pyrroline	Arikit et al. 2011 [260]
Phospholipase D, SPLDα	β-conglycinin α′ subunit	Biolistics	Seed	Changes in phospholipid and triacylglycerol composition	Lee et al. 2011 [261]

Later, in the same year Smith and co-workers further improved this technology by replacing a random fragment of DNA with an intron to link the inverted repeat sequences and formed a self-complementary hairpin structure, also referred to as intron-containing self-complementary hairpin RNA (ihp) [31]. Wesley and co-workers in their study proposed that using an intron-derived spacer sequence enhances the silencing efficiency by improving the stability of ihp RNA [32]. ihp induced RNAi can silence a large number of genes with a remarkable 100% transgene silencing efficiency and no visible phenotypic change [263-265]. In plants, the majority of transgene-induced RNAi silencing is achieved using ihp [185]. Therefore, in our current study, we designed an *IPK2* specific ihp construct to target PA reduction in soybean.

Many RNAi cloning vectors are available commercially. However, care must be taken to avoid vectors designed to express dsRNA/ihp constitutively, since, their ectopic expression may lead to undesirable effects on normal plant phenotype. Vectors designed to include tissue or organ-specific promoters are in the majority of cases a better option [33-35]. Because soybean seeds are the most valued product of soybean plants, one of the main focus of RNA silencing in soybean has been on modification of seed components. In our current study, we used vicilin storage protein gene promoter to express the designed ihp for down-regulating PA only in the soybean seeds.

Furthermore, a reliable protocol for stable transformation of soybean explants with RNAi silencing constructs using *Agrobacterium* as the mode of transfer is still unavailable. The general procedure for generating dsRNA encoding plants involves: 1) constructing an RNAi vector corresponding to the target gene along with, a specific promoter to drive its expression at the desired location and, a marker gene to trace its

existence in the host, 2) transferring the silencing cassette into a binary vector, 3) mobilizing the binary plasmid into *Agrobacterium*, and 4) *Agrobacterium* transformation of the host plant. The success of the process is dependent on several factors primarily, bacterial strain interaction with genotype, choice of the selectable marker gene, type of explants, participation of transformed cells in de novo shoot formation, etc. [266]. Thus, to generate strong, consistent and heritable transformation events more knowledge and effort is required for optimizing the various parameters involved. The type of explant and choice of the selectable marker are two critical factors which determine the success of a transformation experiment. Therefore, in our study, we used a cotyledonary node method to transform the silencing construct into soybean which also bears an herbicide-resistant *bar* gene as the plant selectable marker. Regeneration of fertile soybean transgenics from cotyledonary node explants saves time and effort as it does not require the maintenance of any parent or long-term cultures [267]. The *bar* gene used for selecting transgenic plants was originally cloned from a bacterial species *Streptomyces hygroscopicus* which produces the tripeptide bialaphos as a secondary metabolite. Bialaphos also commercially available as Basta, Buster, or Liberty contains phosphinothricin, is a glutamate analogue which inhibits glutamine synthetase enzyme. *Bar* gene encodes an acetylating enzyme phosphino-thricin acetyl transferase (PAT) which inhibits the herbicide bialaphos by causing acetylation of phosphinothricin and thus confer resistance against it. Non-transgenic plants devoid of *bar* gene show lethal inhibition of glutamine synthetase enzyme which results in toxic accumulation of ammonia in both bacterial as well as the plant cells. Ammonia accumulation kills bacteria and causes severe damage to the plant tissues which ultimately result in the death of the plants treated with bialaphos. *Bar* gene allows stringent selection of only bialaphos-resistant transgenic plants. It has

been successfully used as an important marker in the selection of various transgenic plants including tomato and potato [268], rice [269], wheat [270], sorghum [271, 272], etc.

2.9 Advantages of RNA interference

Gene silencing by RNAi technology provides several advantages. First, since RNAi is dominant in nature, it is comparatively easy to achieve genetic manipulation and subsequently screen the transgenic individuals. Second, the expression of genes can be manipulated spatially and/or temporally by using tissue/organ-specific promoters. Third, its sequence-specific mechanism results in a higher gene silencing efficiency and potency. Fourth, it can silence an endogenous target gene without causing any modification in its structure or function and thus evade lethality. Fifth, it can overcome potential gene redundancy problems by simultaneously inhibiting the expression of several homologous genes.

Chapter 3

Isolation & characterization of soybean *inositol polyphosphate 6-/3-/5-kinase*, a key phytic acid pathway gene by *in-silico* analysis, homology modeling, & docking simulation

3.1 Introduction

Phytic acid is a principal reserve form of phosphorus and inositol which accumulate in plant seeds to feed the developing embryo and guarantee seed germination and plantlet emergence. However, during its accumulation PA forms mixed salts known as phytin with various minerals such as Fe, Ca, Mg, K, Mn, and Zn which renders it indigestible by non-ruminants and thus compromise their mineral absorption [98, 273]. This undigested phytate is excreted in feces and later digested by microbial phytases present in the soil. Phosphorus and other minerals released from PA digestion contribute to environmental pollution by causing eutrophication of water bodies [98]. Collectively, these problems have developed a significant interest in altering seed phytate production in agriculturally significant plants.

PA biosynthesis in plants is proposed to occur through the lipid-independent pathway [19, 274] and the lipid-dependent pathway [275, 276] which is present in most eukaryotic cells. The two routes differ only in the steps leading to the synthesis of inositol trisphosphates (InsP3) [15]. In lipid-dependent pathway, *myo*-inositol is sequentially converted to PtdIns, PtdIns(4)P1 and PtdIns(4,5)P2 by phosphatidyl-inositol synthase, phosphatidylinositol 4-kinase and phosphatidylinositol 4,5-bisphos-phate kinase respectively. PtdIns(4,5)P2 is finally hydrolysed by phospholipase C to

yield Ins(1,4,5)P3. On the other hand, lipid-independent pathway involves sequential phosphorylation of *myo*-inositol to soluble inositol phosphates (IPs). In both the pathways, phosphorylation from InsP3 to PA uses three main classes of inositol phosphate kinases viz. inositol polyphosphate 6-/3-/5-kinase (IPK2), inositol 1,3,4-trisphosphate 5-/6-kinase (ITPK) and inositol polyphosphate 2-kinase (IPK1). The first step of InsP3 to PA conversion is catalyzed by IPK2 enzyme, also known as inositol polyphosphate multikinase. It is a promiscuous enzyme which phosphorylates a number of InsP3/InsP4/InsP5 substrates [277] and thus plays a key role in regulating these InsPs as well as PA turnover. Therefore, if we perturb PA synthesis at this particular step, we can achieve a desirable reduction in PA levels. In order to achieve this, it is important to first lay a groundwork by characterizing the gene encoding this enzyme that will further act as a building block. We thus, conducted an *in-silico* research to study the enzyme molecule in detail as well as investigated its spatio-temporal expression patterns in soybean.

3.2 Materials and Methods

3.2.1 Plant materials

Soybeans [*Glycine max* (L.) Merr.] cv. Pusa 16 were procured from the fields of Genetics Division at IARI, Delhi, for use in this study. For spatial and temporal expression profiling, stem, flower, root and leaf tissues were harvested from plants which were 30 days old while developing seeds were harvested regularly after flowering until maturation and sorted based on their sizes. All the samples were frozen immediately in liquid nitrogen and stored at -80°C until further use.

3.2.2 Isolation and cloning of *GmIPK2*

3.2.2.1 RNA extraction and cDNA synthesis

Total RNA was extracted from seeds, stems, flowers, roots, and leaves using TRIzol reagent method (Invitrogen, Thermo Fisher Scientific). All the reagents used for RNA extraction were prepared in diethylpyrocarbonate (DEPC) treated water and glassware were baked at 180°C for 5-8 hours [278]. 50-100 mg of all the samples were homogenized in liquid nitrogen and extracted at room temperature with 1 ml of TRIzol reagent and 200 µl chloroform by vigorous stirring to form an emulsion followed by incubation for 5 min. The extracts were centrifuged at 12,000 rpm, 4°C for 20 min. Total RNA was recovered from the aqueous phase by extracting with one volume of chilled isopropanol and subsequent precipitation by centrifugation at 12,000 rpm for 20 min following incubation at room temperature for 10-15 min. The pellets thus obtained were washed twice with 70% ethanol and air dried before finally suspending in RNase-free water. The integrity of isolated RNA was evaluated by denaturing agarose gel electrophoresis (see Appendix II), and its yield was determined using NanoDrop 2000 spectrophotometer (Thermo Scientific, USA). The total RNA was then prepared for cDNA synthesis by treating with DNase I (NEB, UK) to remove any genomic DNA residues.

Single-stranded complementary DNA (sscDNA) synthesis was subsequently performed from the DNA-free total RNA using RevertAid^{TM} H Minus Reverse Transcription kit (Thermo Scientific, USA). One microgram of total RNA was transcribed by preparing the cDNA synthesis mix as follows:

cDNA synthesis mixture

Reagent	Volume (µl)
Total RNA	1 µg
Oligo dT Primer	1
5X Reaction Buffer	4
RiboLock RNase Inhibitor (20 U/µL)	1
10 mM dNTP Mix	2
RevertAid M-MuLV RT (200 U/µL)	1
NFW	to 20

The reaction mix was incubated for 1 hr at 42°C, after which the reaction was stopped by inactivating the enzyme at 70°C for 5 min. The cDNA product thus obtained was placed at -80°C for prolonged storage.

3.2.2.2 PCR amplification, cloning, and sequencing

Oligonucleotide primers used for the amplification of *GmIPK2* cDNA were designed using soybean *IPK2* transcript sequence (Glyma.12G240900.1) retrieved from the plant genomics resource Phytozome v9.1. The primers were designed to isolate complete *GmIPK2* coding sequence (CDS) as well as a fragment of its 3' untranslated region (UTR) with the aim to design RNAi silencing construct in the future specific to this highly conserved region. The primers thus designed were tested using a primer design and analysis tool OligoAnalyzer (IDT, USA). PCR amplification of *GmIPK2* was subsequently achieved using these upstream and downstream primers, by preparing the reaction mixture and following the generalized cycling conditions described as follows:

Primers sequences used for amplification:

GmIPK2 FP: -5′ ATGCTCAAGATCCCGGAG 3′

GmIPK2 RP: -5′ CAGTTAGTCTGCGACACTAATTCAAGC 3′

PCR reaction mixture

Reagent	Volume (µl)
10X PCR buffer	2.5
dNTP mix (10mM each)	1
25mM $MgCl_2$	2
GmIPK2 FP (200 pmol/µl)	0.5
GmIPK2 RP (200 pmol/µl)	0.5
Template DNA (50ng/µl)	2
TaqDNA polymerase (5U/µl)	0.5
NFW	to 25

Once prepared, the PCR reaction mixture was placed on a PTC-100^{TM} Peltier thermal cycler (MJ Research) to perform *GmIPK2* amplification.

PCR cycling protocol

Steps	Temperature	Time	Cycles
Initial denaturation	94^{o}C	4 min	1
Denaturation	94^{o}C	30 sec	35
Annealing	62.5^{o}C	30 sec	
Extension	72^{o}C	30 sec	
Final extension	72^{o}C	10 min	1

On completion, the quality and the approximate size of the amplification product was evaluated by agarose gel electrophoresis (see Appendix II) alongside a 1 kb DNA ladder. The PCR reaction was gel-purified by using gel/PCR DNA Fragments Extraction Kit (DF100) from Geneaid to remove any undesired amplification products as well as primer dimers which subsequently improve the ligation efficiency. The desired amplicon was then ligated in the TA cloning vector pGEM-T Easy (Promega, USA) by following the manufacturer's instructions as described below:

Ligation mixture

Reagent	Volume (µl)
2X Rapid Ligation Buffer	5
pGEM-T Easy Vector (50 ng/µl)	1
GmIPK2 insert (4 ng/µl)	2
T4 DNA ligase (3 Weiss units/µl)	1
NFW	to 10

The ligation mixture was incubated at 4^{o}C overnight and used directly for transformation of competent *E. coli* DH5α cells (see Appendix II). The recombinant plasmids were identified by blue-white screening, isolated (see Appendix II), and confirmed by restriction digestion with *Eco*RI. The nucleotide sequence of *GmIPK2* was determined on ABI 3730xl automated DNA Analyzer by analysis of overlapping plasmid clones using T7/SP6 universal primers. The sequence so obtained was deposited in INSDC database GenBank.

3.2.3 Gene expression analysis by RT-PCR and qRT-PCR

To study the spatio-temporal expression profile of *GmIPK2*, we first performed semi-quantitative reverse transcription PCR (RT-PCR) to obtain an expression pattern and then further estimated the transcript levels by relative quantitative RT-PCR also known as real-time PCR or qRT-PCR. RT-PCR was performed with first strand cDNA template, using the same pair of *GmIPK2* specific primers and thermal cycling conditions which were described for its cloning. We further performed qRT-PCR analysis on a PikoReal^{TM} 96 platform (ThermoFisher Scientific, USA). The reactions were set up using DyNAmo Flash SYBR Green qPCR Kit (Thermo Scientific, USA) with cDNA first strands as the template DNA to monitor the quantitative amplification of *GmIPK2*. The expression of *GmIPK2* was normalized to an endogenous control, the housekeeping gene phosphoenolpyruvate carboxylase (*PEPCo*) [279, 280]. The

primers for both the gene and the internal control were designed using PrimerQuest tool by IDT, USA. The experimental reactions were prepared by adding components in the following order:

Primers sequences used for amplification:

q*IPK2* FP: -5′ CGCGGATCCGCGTTGCAGAAGCTCAAG 3′

q*IPK2* RP: -5′ TCCCCGCGGGGAGCGACACTAATTCAAG 3′

q*PEPCo* FP: -5′ CATGCACCAAAGGGTGTTTT 3′

q*PEPCo* RP: -5′ TTTTGCGGCAGCTATCTCTC 3′

qPCR reaction mixture

Reagent	Volume (µl)
2X Master mix	10
Primer Mix (0.25 µM each)	2
cDNA template (5 ng/µl)	4
NFW	to 20

The qPCR mixture was mixed thoroughly to ensure homogeneity and dispensed into 96 wells plate. The reactions were placed in the instrument, and the following program was run:

qPCR cycling protocol

Steps	Temperature	Time	Cycles
Initial denaturation	95^{o}C	4 min	1
Denaturation	95^{o}C	15 sec	40
Annealing/extension	60^{o}C	30 sec	
Data acquisition	Fluorescence data collection	-	
Melt curve	60-95^{o}C	-	1

The cycler was set to collect fluorescence data during the annealing/extension step of each cycle. A melting curve was also recorded post-amplification. Three technical replicates were carried out for each of the three biological replicates to minimize variation in the output. The $2^{-\Delta\Delta CT}$ method described by Livak and Schmittgen [281]

was employed to calculate the relative abundance of *GmIPK2* transcripts.

3.2.4 Sequence analysis and phylogenetic tree construction

We performed a homology search of the deduced GmIPK2 sequence using the PSI-BLAST tool of NCBI, to shortlist homologous IPK2 sequences based on their percentage sequence identity. The primary sequence composition of shortlisted sequences was determined using EMBOSS Pepstats [282]. The physico-chemical parameters of proteins were predicted using the ExPASy web server tool ProtParam [283]. The subcellular location was predicted using TargetP 1.1 Server with 95% specificity cut-offs [284] and the outcome was compared to predictions obtained from MemType-2L [285], SubLoc v1.0 [286], WoLF PSORT [287], and CELLO v2.5 [288]. Its transmembrane topology was predicted with PSIpred [289], TMpred [290], and NPS@ web programs [291]. Presence of potential secretory signal peptides or mitochondrial targeting peptides was analyzed with SignalP 4.1 web server [292]. M-Coffee multiple sequence alignment (MSA) of the selected amino acid sequences was carried out to produce quality alignments which served as the basis for phylogenetic analysis to detect its evolutionary placement and phylogenetic similarity with other similar genes [293]. The evolutionary tree was constructed using the Neighbor-Joining (NJ) clustering method and Poisson model for computing corrected amino acid distances [294] in MEGA (Molecular Evolutionary Genetic Analysis) Version 6.0 [295]. Bootstrap replications were set at 1000 to assess the degree of confidence for each clade of the observed tree. All the branches generated in < 50% bootstrap replications were collapsed. The final image was rendered with the Interactive Tree of Life server (iTOL) [296].

3.2.5. Promoter isolation and prediction of regulatory motifs

A motif search was carried out in the promoter sequences using PlantCARE [297] to define putative *cis*-acting elements involved in the regulation of *IPK2* expression. PlantCARE is a database of known *cis*-regulatory elements present in plants as well as a web-tool which enables *in silico* analysis of promoter regions. To identify these cis-regulatory elements, we retrieved around 2 kb upstream sequence of the IPK2 homologs by NCBI's nucleotide BLAST program and fed them to the PlantCARE web tool. Its database identified individual sites and consensus sequences of regulatory elements as well as defined their properties based on the existing literature. Unlike CDS, these regulatory sequences indirectly influence their immediate phenotype.

3.2.6 Secondary structure analysis and domain prediction

Multiple sequence alignment (MSA) of the selected amino acid sequences was carried out by M-Coffee alignment, and the final alignment was curated in Jalview [298]. The conserved amino acid sequences so identified were highlighted in blue. Ungapped motifs were also detected using MEME web tool available on MEME suite 4.11.1 [299]. The motifs present were further verified using My Hits motif scan tool [300]. The domain composition was analyzed using CDD tool (www.ncbi.nlm.nih.gov/ cdd) on the NCBI server. A combined secondary structure consensus from amino acid sequences was built based on predictions made with SOPMA (nearest-neighbor method) and PHD (neural networks method). This helps in correct prediction of 82.2% of the amino acid residues [301, 302]. Cysteine species and disulfide connectivity of protein sequences were determined using web tool DiANNA [303]. A

secondary structure topology map of the 3D model was built with ProMotif [304] and Topdraw [305].

3.2.7 Protein modeling and model quality assessment

The three-dimensional (3D) model of GmIPK2 was constructed by homology modeling using automated comparative protein modeling servers, SWISS-MODEL [306] and PHYRE2 [307] as well as a standalone comparative modelling program MODELLER 9.16 [308]. Comparative modeling of protein 3D structures usually consists of five steps. The first two steps deal with the search and selection of related protein structures for use as an appropriate template. We obtained the template for homology modelling by alignment of the *in silico* translated amino acid sequence of *GmIPK2* against the PDB database of NCBI by PSI-BLAST search. Protein with the best hit having the lowest E-value, highest sequence similarity was selected as the most appropriate template. The next task in modeling is to generate an accurate target (GmIPK2 in our case) and template sequence alignment. The accuracy of the alignment is of great importance since the quality of the model is dependent on it. In MODELLER, target-template sequence alignment is done using a dynamic align2d() algorithm which considers structural details of the template for generating alignment and thus reduce chances of errors by one third in comparison to the standard alignment procedure. All the parameter files were prepared in PIR format readable by MODELLER. The script files used to run MODELLER are written in the Python programming language. A 3D model of the target was constructed, once a good alignment between the target and template was generated. The construction is a step-wise process that begins with modeling the core regions. The backbone conformation from the template is transferred to the target, and the residues that are conserved retain

these conformations while the non-conserved side chains are transferred without any reference to the template. The next step is modeling of loop regions which is a challenging step since loops tend to be less conserved and of varying lengths in the target and the template. A loop homology search is generally performed where the sequence of the loop is used to search for loops with similar sequence and known conformation to construct a non-conserved loop region. The last step is the rotation of non-conserved side chains and optimization [309]. MODELLER calculates the 3D model using its automodel class which relies on the satisfaction of spatial restraints. There are two main sources of spatial restraints: 1) restraints applied on the dihedral angles and the distances of the target protein following its alignment, and 2) restraints on the stereochemical properties of the target protein due to the applied Charmm-22 force field. By default, five comparative models are generated, and each model is assessed for quality based on its molpdf, GA341 [310], and DOPE (Discrete Optimized Protein Energy) [311] scores. The Molpdf scoring function is derived from collective restraints, GA341 is based on statistical potentials and DOPE is based on the atomic distances. A model with the least molpdf and DOPE scores was selected as the final model, and its quality was further assessed. After a model is built, it is important to check it for possible errors. In the final step of comparative modelling, the stereochemical and energetic properties of the obtained models were evaluated with RAMPAGE [312], VERIFY 3D [313], and ProSA servers [314]. The quality of models was additionally assessed by comparing it with the template structure on the UCSF Chimera's MatchMaker tool [315] and calculating the Cα root mean square deviation scores (RMSDs) for each of the comparative models. The secondary structures of the final and template proteins were also compared by pairwise 3D alignment using MATRAS 1.2 [316]. Also, any disordered regions in the model were

predicted using PrDOS server [317]. The final model was chosen based on the majority in the best scores from all the quality analyses. In order to make an accurate model as much possible with the available structural data, careful thought was put in throughout the modeling process, including literature, database searches, multiple sequence alignments as well as model building. The final model was used for further molecular dynamics and molecular docking analysis.

3.2.8 Refinement of homology model

We subsequently optimized the obtained GmIPK2 model by performing molecular dynamics (MD) simulation. The process was conducted in explicit solvent for 50 ns using the Groningen Machine for Chemical Simulations, GROMACS 4.5.5 package [318]. The whole simulation was divided into the following major steps:

3.2.8.1 Energy Minimization: At the beginning of a simulation, each atom of a molecular system strives to find a stable arrangement in the space in response to an applied molecular mechanics force field. To begin simulation, we solvated our protein model by adding simple point charge (SPC216) water in a 0.7 nm pre-equilibrated cubic box. We then applied a GROMOS96 53A6 force field recommended for simulation of biomolecules in explicit water in order to begin energy minimization. During minimization, the force experienced by each atom is reduced in a stepwise fashion which is reflected in a change in its geometry until the net force experienced by each atom is close to zero and a global energy minimum is attained. The process is divided into discrete time steps, and the force acting on each atom is computed.

We used a maximum 1000 steps to perform minimization of our protein. In every step, the position and velocity corresponding to each atom were calculated using Newton’s laws of motion. There are chiefly three procedures to achieve minimization

viz. steepest descent, conjugate gradient, and Newton-Raphson. We used the steepest descent algorithm to perform the energy optimization of our protein. In steepest descent, we move along the negative of the gradient direction such that we get a maximum decrease in the potential energy to attain a global minimum. This procedure requires the smallest amount of CPU time per step of minimization.

3.2.8.2 Equilibration: In this stage of MD, we performed equilibration of solvent molecules and ions around the protein molecule. This process is divided into two phases. The first phase of temperature equilibration performed using NVT ensemble (constant Number of particles, Volume, and Temperature), also referred to as "isothermal-isochoric" or "canonical" where the temperature of the system is raised to the desired temperature of simulation to establish the proper orientation of the protein about the solute. In this case, we heated the system to 300K and stabilized it for 100 ps. Once we arrived at the correct temperature, we applied pressure to the system until it reached a proper density. To achieve pressure equilibration, we engaged the NPT ensemble (constant Number of particles, Pressure, and Temperature), also referred to as "isothermal-isobaric" ensemble. A stable density of 1008 kg m-3 corresponding to that of the SPC model over time indicate satisfactory equilibration of the system, and thus we can move to the next stage of simulation.

3.2.8.3 Production MD: In this stage, known as the dynamic stage we release the applied position restraints and run the simulation. The only things that we changed were that we removed the position restraints, turned off the pressure coupling and increased the simulation length to 50 ns. We performed the MD simulation just like in the previous steps. The entire simulation was done using supercomputing facility at IIT Delhi on remote servers linked from windows desktops with ssh putty, and all the file were transferred using the WinSCP tool.

3.2.8.4 Analysis: After successfully completing the production run, we analyzed the results by collecting data using a post-processing tool, trjconv which extracts the coordinates and fix for periodicity. We used this corrected trajectory to study the stability and function (i.e. dynamics) of our simulated protein structure by studying how the distance between the Cα atoms of the protein residues evolves with time. For this, we calculated RMSD and root mean square fluctuations (RMSF) of Cα backbone atoms of the model along the simulation time using *g_rms* and *g_rmsf* tools and analyzed the output using simple plotting program called xmgrace [319].

3.2.9 Molecular docking and MD simulation of the docked complexes

Molecular docking was performed to predict and characterize the 3D structure of the protein-ligand complexes which will help in determining the crucial amino acid residues involved in complex formation. Based on the available literature for IPK2, D-*myo*-inositol 1,4,5-trisphosphate (PDB: 5GUG-I3P) & D-*myo*-inositol 1,4,5,6-tetra-*kis*phosphate (PDB: 4A69-I0P) ligands were selected for optimized GmIPK2 protein receptor (Figure 3.1).

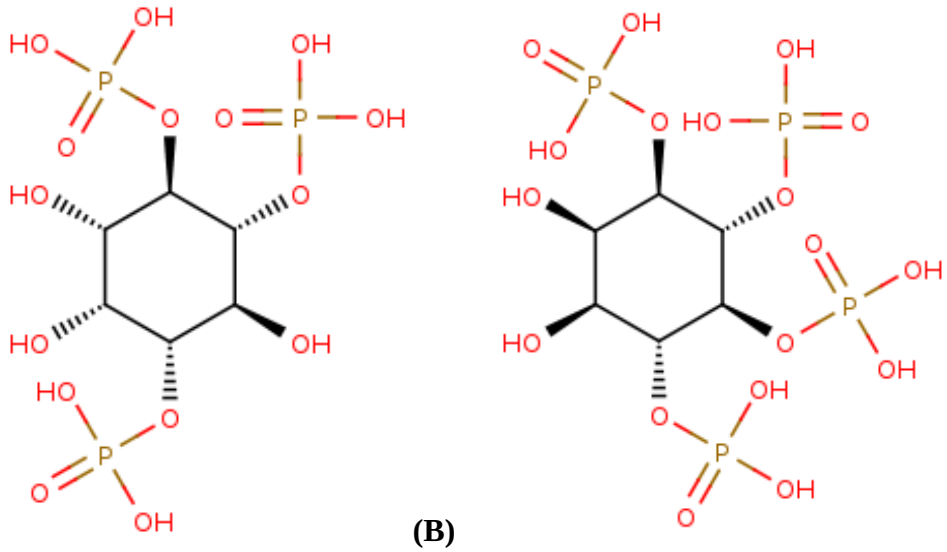


Figure 3.1: Chemical structure of ligands: (A) D-*myo*-inositol-1,4,5-trisphosphate (PDB ID: 5GUG-I3P) (B) D-*myo*-inositol-1,4,5,6-tetra*kis*phosphate (PDB ID: 4A69-I0P).

Docking of these ligands onto the GmIPK2 protein was performed using AutoDock Vina (version 1.1.2). The process of molecular docking consists of two parts: 1) sampling to generate trial protein-ligand poses and 2) scoring to evaluate a pose by assigning it a value that presumably reflects its accuracy. AutoDock Vina uses a sophisticated sampling algorithm, Iterated Local Search global optimizer to generate a diverse set of conformations of the compounds *in vacuo* by the sampling of torsional and rotational degrees of freedom [320], and further, evaluate these generated poses using its inbuilt scoring function. These conformations are placed into the binding pocket of the homology model as a starting point for global optimization of the energy function where iterations of random torsional and positional moves are performed, followed by local energy minimization. The conformations of the compounds are either accepted or rejected based on energy [321]. The low energy conformations are stacked, saved, and ranked based on the docking energy. The quality of docking depends on the flexibility and quality of the protein and ligand structures and, thus important to consider. Even though the fully flexible docking approach produces highly accurate results, depicts the real-life nature of the complex, it is often ignored as it is computationally very demanding. We thus used a more common approach of semi-flexible docking where the ligand is treated as flexible and the target as rigid [322]. AutoDock Vina reads all the molecules in a simplified PDB file representation, termed PDBQT and thus coordinates of both the GmIPK2 protein and its ligands were prepared using MGL. The solvent atoms and water molecules were removed, and the polar hydrogen atoms were added to the protein. Only the ligand molecules were allowed to rotate and explore the binding pockets. Vina uses a customized rectangular 3D cartesian grid for defining the binding site of the protein and for efficient geometric scoring. We customized the dimensions of the grid to make sure that the

size of the search space is large enough for proper rotation of the ligand. The program was then run to generate ligand poses, each with definite binding energy (kcal/mol) which is calculated based on Vina's default scoring function. The scoring function ranks the possible conformations/orientations of the ligands according to binding tightness in the pocket, ideally giving top rank to the pose closest to the experimentally determined mode. The conformations with the lowest binding affinity were chosen, and the interaction diagrams were generated using Discovery Studio Visualizer 4.1 (Accelrys Software Inc., USA, 2013). The amino acid residues that were present at a distance of approximately 2Å were considered as the binding partners of the ligands. The active amino acid residues were also predicted by combining results of three different interface prediction web servers, CASTp [323], FTSite [324], and FunFOLD2 [325] into a consensus. The final complexes were equilibrated by MD simulation by following the procedure described for protein model in explicit water under the section 3.2.7 but by applying restraint on the ligands to prevent them from moving away from the binding site. Once equilibrated, we simulated the complexes for 50 ns and analyzed their resulting trajectories to understand the relative stability of the ligands inside its binding pocket using *g_rms*, *g_rmsf*, and *g_mmpbsa* tools. MD simulations based on molecular mechanics and Newton's laws of motion [326] are often considered as an integral part of docking; however, they differ in their basic principles.

3.3 Results and Discussion

3.3.1 Isolation, cloning, and sequencing

Isolation of a gene is the pre-requisite for its thorough characterization. Generally, to

isolate any gene of interest there are two major approaches: 1) PCR amplification, and 2) Genomic/cDNA library screening. In the present study, we used a PCR based approach to fish out the PA biosynthetic pathway gene, *GmIPK2* from *G. max*. We first performed a BLASTN analysis in Phytozome v9.1 [327] using the soybean *IPK2* gene sequence available in NCBI (GenBank: NM_001250522) as the query to retrieve a 1241 bp transcript sequence. Based on the sequence data thus derived, we designed primers specific to amplify complete *GmIPK2* coding sequence (CDS) as well as a fragment of its 3' untranslated region (UTR) with the aim to design silencing construct in the future specific to this highly conserved region because of its substantial role in gene regulation. To accomplish this, we isolated total RNA from 8-10 mm developing seed stage (Figure 3.2A), synthesized cDNA from it and performed PCR amplification (Figure 3.2B) from the synthesized cDNA template as per the protocol described in section 3.2.2. We then cloned the amplicon into the pGEM-T easy vector (Figure 3.3A) and introduced it into the bacterial host *E. coli* (DH5α). We further verified the putative cDNA clones by restriction analysis with *Eco*RI and sequence characterized to ~910 bp residues in length (Figure 3.3B).

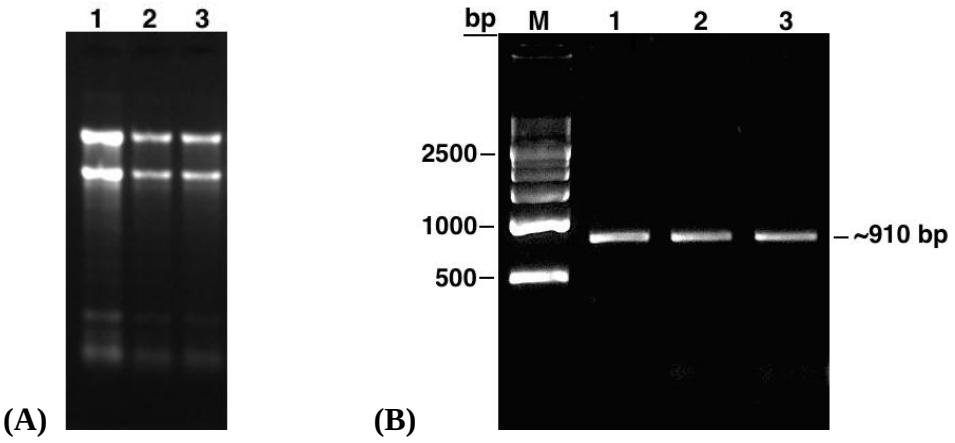
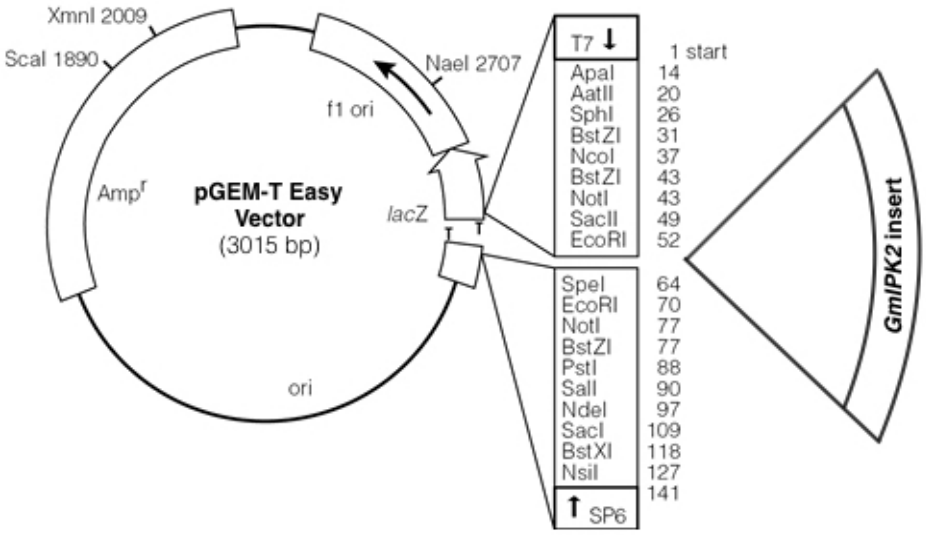
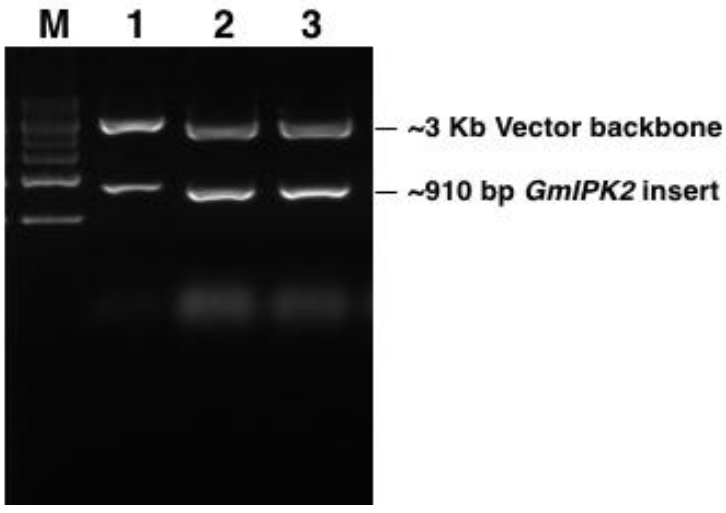


Figure 3.2: (A) Electrophoresis of total RNA isolated from developing seeds, 8-10 mm in size of *G. max* on 1% agarose gel (Lanes 1-3). **(B)** PCR amplification product

of *GmIPK2* gene separated on 0.8% agarose gel. Lane M: 500 bp DNA ladder, lanes 1-3: ~ 910 bp amplified product.



(A)



(B)

Figure 3.3: (A) Simplified map of ~910 bp *GmIPK2* cDNA cloned into pGEM-T Easy vector by TA cloning. **(B)** Restriction digestion of the clones with *Eco*RI. Lanes 1-3 showing release of vector backbone (~3 kb) and *GmIPK2* insert (~910 bp), Lane M: 500 bp DNA ladder.

>KF297702.1 Glycine max cultivar Pusa 16 inositol polyphosphate 6-/3-/5-kinase mRNA, complete cds

ATGCTCAATATCCCGGAGCACCAGGTGGCCGGGCACAAGGCCAAGGACGGGATCCTGGGCCCGCT
CGTCGACGATTTTGGAAAATTCTACAAGCCCCTCCAGACCAACAAAGACGACGACACCCGCGGTT
CCACCGAACTCTCCTTTTACACCTCTCTCGCCGCCGCCGCCCACGACTACTCCATCCGCTCCTTC
TTCCCCGCCTTTCACGGCACCCGCCTCCTGGACGCCTCCGACGGCTCCGGTCCCCACCCTCACCT
GGTCCTGGAGGACCTCCTCTGCGGCTACTCCAAACCCTCCGTCATGGACGTAAAGATCGGCTCCA
GAACCTGGCACCTGGGAGACTCCGAGGACTACATCTGCAAGTGCTTGAAGAAGGACAGAGAGTCC
TCTAGCTTGCCCTTGGGTTTCAAAATCCCGGGAGTCAAGGACTCTATCTCCTCCTGGGAACCTAC
CAGGAAATCTCTCCAGTGTCTATCCGCCCATGGTGTTGCACTTGTTCTCAACAAGTTCGTTTCCT
CTAATAATATCAACCATGATGATCATCATCCCGATTGCGCTTTCGCAACGGAGGTCTACGGCGCC
GTTTTGGAGCGCTTGCAGAAGCTCAAGGACTGGTTCGAGGTTCAGACGGTGTATCACTTCTATTC
TTGTTCTGTTCTTGTGGTGTACGAGAAGGATCTAGGGGAAAGGAAAGCTACCAACCCTCTGGTCA
AACTCGTTGACTTTGCACACGTGGTGGACGGAAACGGTGTCATTGATCACAACTTCTTGGGTGGC
CTTTGTTCCTTCATCAAGTTCCTCAAGGATATCCTAGCAGTAGCATGTCTTCACAAGTGA

Figure 3.4: Nucleotide sequence of *GmIPK2* (Accession: KF297702) gene was submitted to GenBank.

It contains a single ORF ~840 bp long which potentially encode a single polypeptide of 279 amino acid residues and a ~70 bp 3′ UTR fragment. We submitted the obtained CDS data to NCBI (GenBank: KF297702) (Figure 3.4) and used the information as a query to conduct protein homology search using the PSI-BLAST algorithm. Amongst sequences producing significant alignment, we shortlisted 30 plant IPK2 sequences besides GmIPK2 based on percentage sequence identity to carry out further *in silico* analysis.

3.3.2 Spatiotemporal expression profiling

Tissue-specific modulation of *IPK2* gene in order to generate a lpa genotype is essential to evade any possible pleiotropic effects. Before this can be achieved it is vital to investigate its spatial expression profile in different tissues as well as its temporal expression profile in developing seeds, to provide an initial point for the strategic achievement of a desired level of silencing.

In order to study the expression of *GmIPK2* gene in different tissues of a soybean

plant and during seed development, we performed semi-quantitative as well as real-time PCR expression analysis using total RNA isolated from root, stem, leaf and flower tissues of 30-day-old *G. max* plants and developing seeds ranging from 0 to 16 mm in size distributed in eight different progressive stages. Semi-quantitative PCR analysis revealed a differential pattern of *GmIPK2* transcript expression across the set of experimental tissues analyzed, with the highest level of transcripts observed in seeds (Figure 3.5A). The same was confirmed by the steady-state qRT-PCR analysis which also detected the highest level of transcripts in seeds (Figure 3.5B).

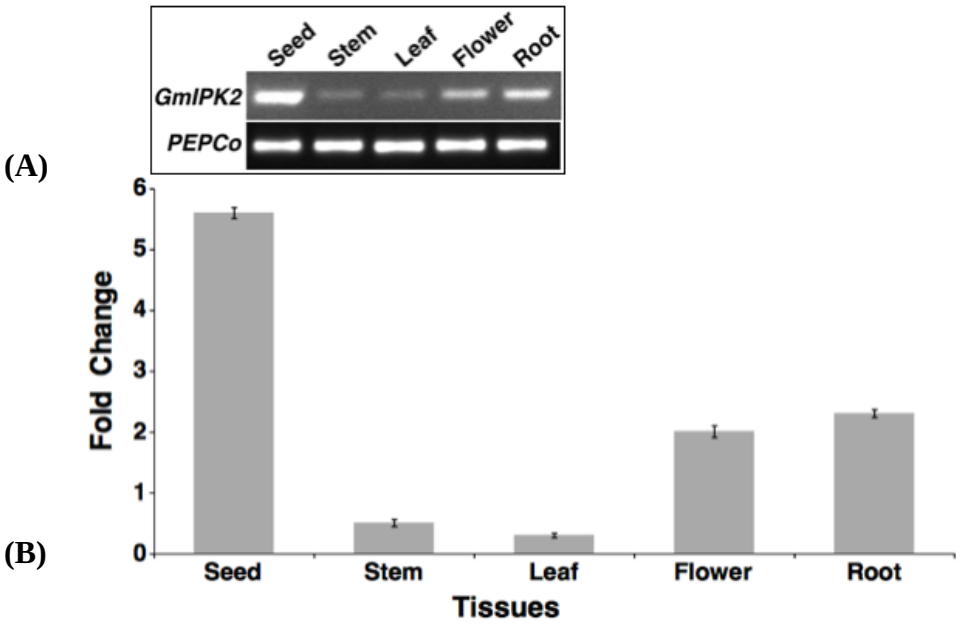


Figure 3.5: (A) RT-PCR expression profile of *GmIPK2* gene in different plant tissues of soybean, separated on 1% agarose gel compared with housekeeping gene *PEPCo* as an internal control. **(B)** Relative quantification of *GmIPK2* transcript levels in the samples analyzed above by qRT-PCR, normalized to soybean housekeeping gene *PEPCo*. The leaf tissue was taken as calibrator. The data is mean of technical triplicates of each of the three biological replicates with error bars indicating standard deviation (SD).

This suggests that *GmIPK2* play a key role in PA biosynthesis in this tissue for use as a primary source of energy during germination. In both the analyses, the same level of

amplification was observed for *PEPCo* housekeeping gene. Since PA is also required for several other vital functions throughout the plant system, a basal level of it is observed in all the other tissues as well. Thus, amongst the tissues analyzed in our study, expression of *GmIPK2* was also recorded in roots and flowers but at a level lower than that observed in the cotyledons. The strong presence of *GmIPK2* transcripts in these tissues may be attributed to its role in regulating cytosolic calcium gradient which correlates with pollen germination, growth of pollen tube [328-331] root growth and root hair development [332-334]. A low level of expression was also observed in stems and leaves as it is studied to be involved in regulating a vital function of axillary shoot branching by participating in auxin signaling [335].

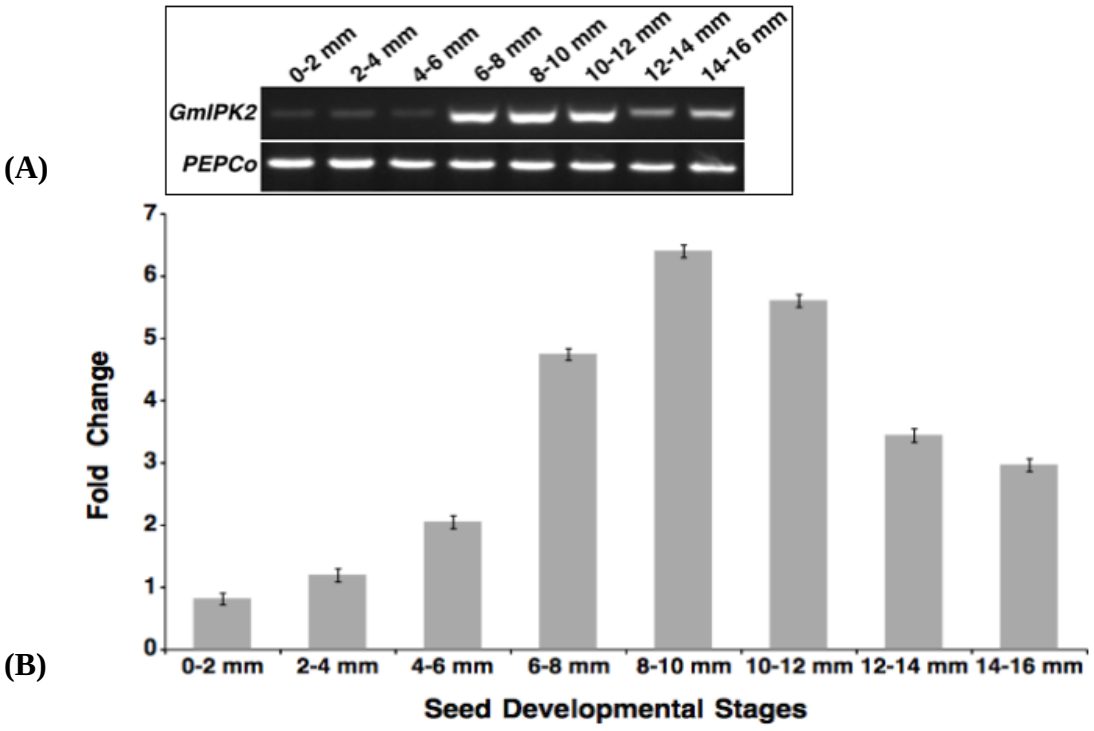


Figure 3.6: (A) RT-PCR expression profile of *GmIPK2* gene in developing seeds of soybean, compared with housekeeping gene *PEPCo* as an internal control. **(B)** Relative quantification of *GmIPK2* transcript levels in the samples analyzed above by qRT-PCR, normalized to soybean housekeeping gene *PEPCo*. The 0-2 mm seed stage was taken as calibrator. The data is mean of technical triplicates of each of the three biological replicates with error bars indicating SD.

We then analyzed its temporal expression pattern during eight progressive seed development stages viz. from 0-2 mm to 14-16 mm. In both semi-quantitative as well as real-time PCR expression analysis, we observed that its expression increased as the development progressed and reached a peak value at the later stages of seed development (Figure 3.6). This pattern of expression coincides evenly with the pattern of accumulation of PA which is linear throughout most of the seed development (Raboy and Dickinson, 1987). We obtained comparative results in the microarray transcriptome study conducted in the past in our laboratory (GEO: GSE69821). This observation can be explained by the increase in production of phosphorus compounds required to support growth and development during initial stages of seed development when the synthesis of the phosphorus reserve, PA is minimal [336].

The spatiotemporal analysis, in summary, identifies seed as the major tissue for its expression with maximum relative expression occurring during the later stages of its development. Thus, from the present study, we can hypothesize that targeting the *GmIPK2* gene expression during late seed development may provide a potential strategy for generating lpa soybean with enhanced nutritional value.

3.3.3 Sequence analysis

3.3.3.1 Identification of regulatory promoter elements

Spatial and temporal gene expression patterns are established and maintained primarily by transcription regulation. The time and location of gene expression are controlled partly by the *cis*-elements present in the promoter regions and their corresponding interaction with transcription factors. Therefore, to understand the transcriptional regulation mechanism of *IPK2* gene, we analyzed the promoter region

of all its homologs and identified different cis-regulatory elements located therein. A 2-kb sequence upstream to the open reading frame was identified and subjected to PlantCARE analysis. Database search revealed the presence of many motifs related to seed-specific promoters, hormone-responsive *cis*-elements (HRE), and *cis*-elements responsive to stresses (DSRE) which together contribute to the differential regulation of our gene (Table 3).

Table 3: Potential *cis*-acting elements identified in the 5′ regulatory sequences of plant *IPK*2s.

Classification	**Name**	**Sequence**	**Source**	**References**
			Organism	
Endosperm	GCN_4 motif	GAAGCCA, TGAGTCA,	*Oryza sativa*	Takaiwa et al., 1996 [337]; Onodera et al., 2001 [338]
	Skn-1 motif	CAAGCCA, TGTGTCA, GTCAT	*Oryza sativa*	Washida et al., 1999 [339]
ABA	ABRE	CACGTG, TACGTG	*Arabidopsis thaliana*	Yamaguchi-Shinozaki et al., 1989 [340]
Gibberellin	GARE	AAACAGA, TCTGTTG, TATCCAC/T	*Brassica oleracea* *Oryza sativa*	Mongkolsiriwatana et al., 2009 [341]; Niu et al., 2016 [342]
Light	Box 4	ATTAAT	*Petroselinum crispum*	Lois et al., 1989 [343]
	CATT-motif	GCATTC	*Zea mays*	Arguello et al., 1996 [344]
	G-Box	CACGTG	*Pisum sativum*, *Arabidopsis thaliana*	Sessa et al., 1995 [345]
	Box-I	TTTCAAA	*Pisum sativum*	Arguello et al., 1996 [344]
	ACE	AAAACGTTTA	*Petroelinum crispum*	Feldbrugge et al., 1996 [346]
	GA-motif	AAGGAAGA	*Glycine max*	Arguello et al., 1996 [344]
MeJA	CGTCA-motif	CGTCA	*Hordeum vulgare*	Kim et al., 1993 [347]; Rouster et al., 1997 [348]
	TGACG-motif	TGACG	*Hordeum vulgare*	
Drought	MBS	T/CAACTG	*Arabidopsis thaliana*	Shinozaki and Yamaguchi-Shinozaki, 2000 [349]; Abe et al., 2003 [350]
Anaerobic	ARE	TGGTTT	*Zea mays*	Nguyen et al. (2003) [351]

Light responsive elements [343-346] were observed most frequently which suggest a probable diurnal regulation of *IPK2* expression. The highly conserved GCN4 and Skn_1 promoter motifs found within the seed storage protein genes of cereals were also located within the *IPK2* promoter. These cis-acting elements play a role in regulating endosperm-specific gene expression [337-339].

We also identified multiple HREs particularly those known to be involved in abscisic acid (ABA) and gibberellic acid (GA3) sensing. An ABA-responsive element (ABRE) with the core sequence PyACGTG/TC and three GA-responsive elements (GARE) AAACAGA, TCTGTTG, and TATCCAC/T [340-342] which correlate with PA accumulation during grain filling [350, 352] were identified. Aggarwal and co-workers [353] reported that *IPK2* is an ABA-induced gene which is antagonistically suppressed by GA3 underlining the crucial role played by these hormones in regulating PA pathway genes. Putative methyl-jasmonate, drought inducibility, and anaerobic induction response elements were also observed which help combat assorted types of abiotic factors that plants are exposed to under natural environment [347-349, 351].

3.3.3.2 Computation of physico-chemical parameters and subcellular localization prediction

3.3.3.2.1 Physico-chemical parameters: GmIPK2 sequence similarity search using PSI-BLAST revealed homology to other IPK2 protein sequences showing maximum similarity with *Glycine soja* (98%) and *Vigna radiata* (74%) (Table 4). Primary sequence analysis of these homologs by ExPASy ProtParam and PEPSTATS tools indicate that leucine is the most abundant aminoacid which makes up to approx. 9 to 11 mole percent of its backbone residues while the percentage of tryptophan and

methionine were found to be the least at approx. 1% across species. The isoelectric points (pI) of all IPK2 proteins were computed to be under 7 suggesting that they are most likely to precipitate in acidic buffers except *Glycine soja, Medicago truncatula, Zea mays, and Sorghum bicolor* which has a pI of 7.1, 8.57, 10.48, and 8.26 respectively indicating their solubility in basic buffers. The calculated pI will be useful for empirical protein purification by isoelectric focusing (IEF) and ion-exchange chromatography. The extinction coefficient (EC) [354] of IPK2 proteins measured at 280 nm in water was found to range from 23950 to 36900 $M^{-1}cm^{-1}$ with respect to the concentration of their aromatic amino acids (11-15%) and cystine (disulfide bonds). These EC values can be used to calculate protein concentration in a solution. The instability indices (Ii) computed for selected IPK2 proteins are used to determine their in vivo half-lives [355]. Rogers and co-workers [356] reported that proteins having Ii values greater than 40 have an in vivo half-life of less than 5 hours, while, those proteins having Ii values less than 40 have a longer in vivo half-life of around 16 hours.

Table 4: Physico-chemical parameters of shortlisted plant IPK2 sequences computed using the ProtParam tool.

Organism	Accession No.	MW	pI	EC	Ii	Ai	GRAVY	-R	+R
Glycine max	AGW99177.1	30979.2	6.29	29910	29.72	89.07	-0.18	35	30
Glycine soja	KHN19419.1	19898.87	7.10	23950	29.73	95.73	-0.059	21	21
Vigna radiata	XP_022634636.1	31482.10	6.59	28420	38.17	93.67	0.045	29	27
Cajanus cajan	XP_020232596.1	29463.71	6.86	28420	33.70	91.31	-0.132	31	30
Phaseolus vulgaris	XP_007133857.1	28951.0	6.59	26930	38.90	90.38	-0.017	27	25
Lotus japonicus	AFK39224.1	31348.76	5.66	32430	42.36	87.74	-0.139	39	30
Cicer arietinum	XP_004510840.1	33061.70	6.12	26930	31.59	91.11	-0.250	38	33
Medicago truncatula	XP_003627882.1	47961.8	8.57	29910	31.89	85.02	-0.344	49	53
Arachis hypogaea	ALT56981.1	32684.01	6.45	28420	36.41	84.01	-0.227	34	32
Arachis duranensis	XP_015937330.1	32883.25	6.45	28420	37.35	84.01	-0.248	35	33
Corchorus capsularis	OMO94774.1	32498.97	5.84	28420	40.75	92.43	-0.159	38	33
Theobroma cacao	EOY22693.1	34213.77	6.32	31400	37.61	84.61	-0.292	39	36

Herrania umbratica	XP_021285810.1	34392.01	6.56	36900	40.38	83.38	-0.307	38	36
Durio zibethinus	XP_022738659.1	32954.54	6.71	28420	38.72	87.83	-0.237	36	35
Solanum lycopersicum	XP_004235863.1	32476.9	5.93	28420	34.32	83.23	-0.261	35	30
Solanum tuberosum	NP_001335929.1	44295.0	5.96	28420	20.21	93.19	-0.254	50	43
Brassica rapa	XP_009112000.1	30640.7	6.25	34380	30.25	80.66	-0.283	36	33
Capsicum baccatum	PHT32205.1	40995.36	5.97	32890	33.47	90.19	-0.297	50	44
Nicotiana attenuata	XP_019263239.1	40865.55	5.51	31400	30.30	93.51	-0.179	48	40
Capsicum annuum	PHT66064.1	41015.41	5.91	32890	32.27	92.01	-0.277	50	43
Capsicum chinense	PHU00941.1	40986.37	5.81	32890	32.27	92.27	-0.271	51	43
Brassica napus	XP_013749343.1	31961.0	5.87	35870	30.23	75.86	-0.387	40	34
Arabidopsis lyrata	XP_020870073.1	33671.08	5.82	31400	29.76	83.47	-0.324	42	35
Prunus avium	XP_021815221.1	31100.28	6.20	31400	26.98	86.41	-0.209	34	30
Trifolium pratense	PNY08557.1	30275.83	5.50	23950	29.99	96.70	-0.054	37	26
Arabidopsis thaliana	NP_200984.1	33486.7	5.72	31400	25.96	80.23	-0.329	40	31
Prunus persica	XP_007209445.1	31303.57	6.50	31400	28.39	85.39	-0.252	35	33
Lepidium latifolium	ACK86969.2	33212.5	6.59	34380	29.25	81.95	-0.285	31	29
Zea mays	XP_008649440.2	26665.45	10.48	29450	65.74	68.79	-0.517	22	37
Aegilops tauschii	XP_020147020.1	30607.86	6.13	26930	38.85	88.43	-0.052	33	29
Sorghum bicolor	XP_002452184.1	34550.64	8.26	28420	39.34	90.93	-0.013	32	34

MW: Molecular weight in g/mol; pI: Isoelectric point; EC: Extinction coefficient in M-1cm-1; Ii: Instability index; Ai: Aliphatic index; GRAVY: Grand average hydropathy; -R: Number of negative residues; +R: Number of positive residues.

Our study showed that Ii values of all the homologs are less than 40 and hence are thermally stable with a long half-life except for *Lotus japonicas* and *Zea mays* kinases that have an Ii above 40 which indicate their possible thermal instability. Thermostability of proteins results from a combination of several factors acting synergistically. Here we assess the thermal stability of our proteins in direct proportionality with their aliphatic index (Ai) which is a relative measure of the volume occupied by the aliphatic side chains [357]. The Ai values determined for IPK2 kinases ranged from 68.79 to 96.70 with those from Brassicaceae family showing lowest thermal stability. This, in turn, is indicative of their greater flexibility at a wide range of temperatures when compared to proteins of other families. Grand average of hydropathicity (GRAVY) number reflects the average hydropathy of a

protein, the positively rated being hydrophobic and negatively rated being hydrophilic in nature [358]. GRAVY index for IPK2 kinases was found ranging from -0.517 to -0.013 which indicate that these proteins interact favourably with water except for *Vigna radiata* which is potentially hydrophobic with an index of 0.045.

3.3.3.2.2 Subcellular localization: Protein localization and target peptide predictions are significant studies as they aid in functional genome annotation. From the acidic amino acid composition of IPK2 homologs determined by the physico-chemical analysis conducted above, we can assume that they are cytoplasmic in nature as opposed to the membrane proteins which have basic amino acid composition for their stability [359]. Further sequence analysis based on TargetP scores (cTP: 0.163, mTP: 0.066, SP: 0.087, other: 0.906) also suggests that IPK2 kinases may be located anywhere in the cell besides chloroplast and mitochondria. Their low signal peptide (SP) score indicate the absence of a signaling pre-sequence which reinforce their soluble nature. Besides, a consensus of predictions obtained from WoLF PSORT, CELLO v2.5, SubLoc v1.0 and MemType-2L servers (Table 5) also established the cytoplasmic character of IPK2 protein.

Table 5: Consensus subcellular localization prediction of GmIPK2.

Tools	Subcellular localization	Value	Reliability index	Method
CELLO v2.5	Cytoplasmic	4	5	SVM
SubLoc v1.0	Cytoplasmic	2.281	Higher Value	SVM
WoLF PSORT	Cytoplasmic	7	14	K nearest neighbors classifier
MemType-2L	Cytoplasmic	Not Defined	Not Defined	Pre-PSS

Pre-PSSM and SVM stands for Pseudo Position-Specific Score Matrix and support vector machines respectively.

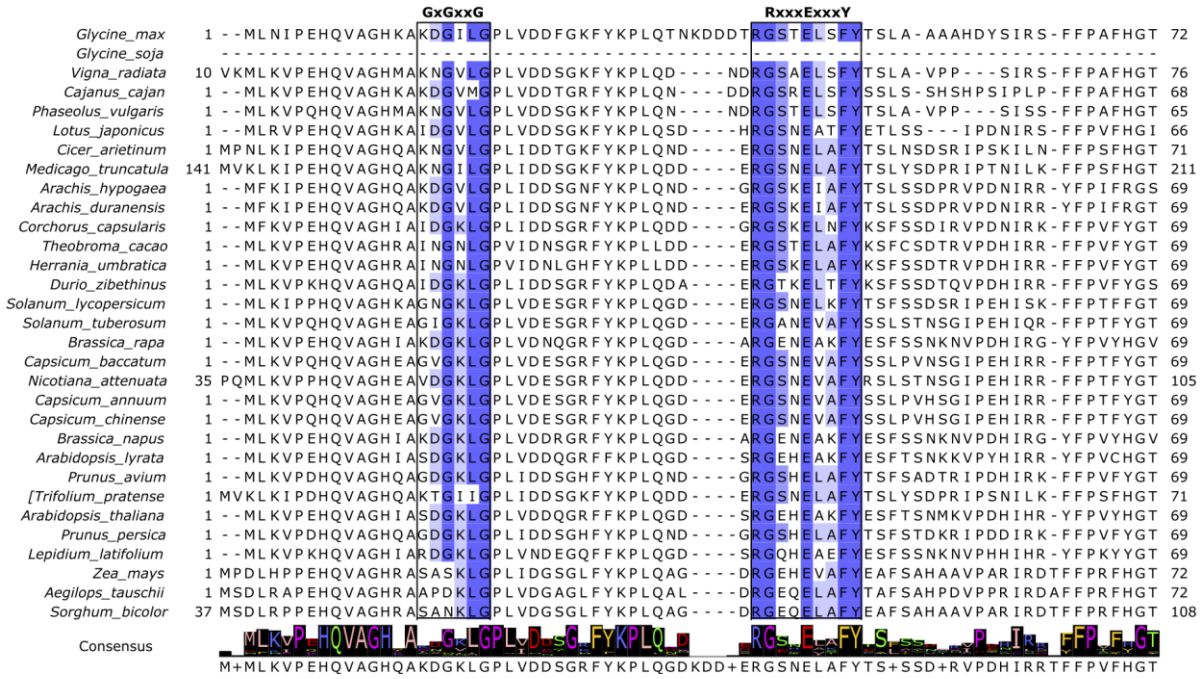
However, TMpred, PSIPRED, NPS@, and DAS servers identified a single consensus C-terminal transmembrane region positioned at 259-274. This prediction could be a mistake based on a sheer coincidence of the presence of a large number of hydro-phobic residues in its C-terminal region and the use of hydrophobicity as the only

criterion to predict membrane-spanning regions. The absence of any transmembrane helix is also well documented in the hydrophobicity plot generated using waveTM server (see Appendix I Figure 1) [360] as the majority of amino acids show negative for hydrophobicity.

3.3.3.3 Domain identification and evolutionary analysis

3.3.3.3.1 Domain identification: Multiple sequence alignment (MSA) of the selected plant IPK2 homologs was performed using M-Coffee web server which is a meta-method for assembling MSA by combining the output of several individual methods into one. The consensus alignment thus generated revealed several significantly conserved motifs and sites unique to inositol phosphate kinases (Figure 3.7). A signature inositol-phosphate-binding motif, PxxxDxKxG was identified in all the aligned sequences [94, 361] which confirms that they belong to IPK superfamily of IP kinases. Holmes and Jogl [362], state that all the members of this superfamily share several strictly conserved signature motifs with each other and are predicted to assume the same overall fold, despite the low sequence conservation. A core catalytic tyrosine kinase motif, RxxxExxxY was also discovered in all the sequences which suggest that they are tyrosine-specific protein kinases [363]. IPK2 sequences from Solanaceae and Rosaceae families were found to contain a Glycine-rich consensus ATP-binding GxGxxG motif characteristic of protein kinase C (PKC) catalytic domain [364]. The classical PKC and plant CDPKs recognized phosphorylation S/TxK/R motif [365-367] was also identified in some of the sequences speculating their role in lipid-dependent PA biosynthetic pathway. Such promiscuous kinase activity suggests that both lipid-dependent and independent pathways regulate PA biosynthesis as well as basic nuclear and cellular processes in plants [108, 368]. A protein recognition LxxLL motif common to all of the aligned sequences indicate

their participation in protein-protein interactions, regulating cell signalling, cell adhesion and transcription [369].



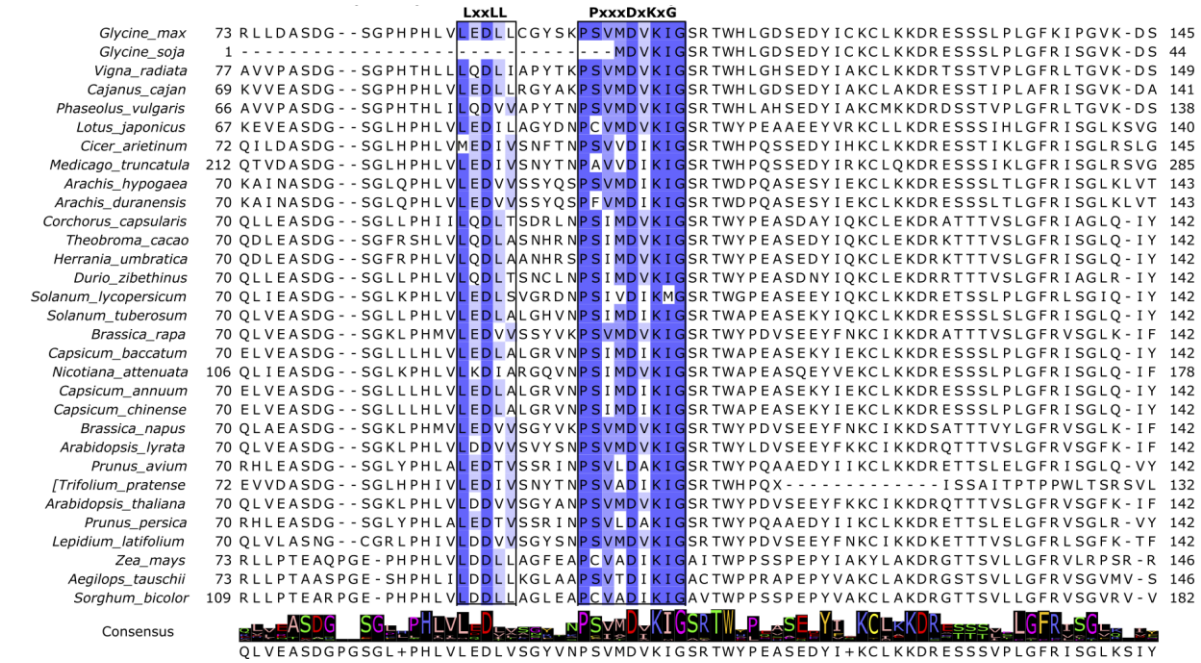


Figure 3.7: M-Coffee *multiple* sequence alignment diagram of selected plant IPK2 protein sequences rendered with Jalview. The sequence motifs shared amongst all the representatives are highlighted in blue colour. The consensus row at the bottom shows the most frequent residue at each column or a '+' if two or more residues are equally abundant.

Further analysis by MEME suite web server identified total eleven conserved ungapped motifs, with motif PxxxDxKxG being the most conserved amongst all IPK2 homologs as indicated by its lowest E-value of 2.9e-487 (see Appendix I Figure 2). The obtained motifs were subjected to further analysis by BLASTP for confirming their annotations which established that they all belong to the IPK superfamily domain (CDD Acc: cl12283) and thus substantiate our previous results. We then explored GmIPK2 protein sequence only by Motif Scan which recognized its diverse protein kinase phosphorylation sites (Table 6). Since phosphorylation acts as a molecular switch in modulating protein function, structural rearrangement, and

cellular localization, we can suggest that GmIPK2 play a critical role in many biological regulatory events of signalling, proliferation, differentiation, and apoptosis. Its role in biological processes as diverse as mRNA export [63], DNA repair [64], chromatin structure regulation [65, 66], plant pathogen resistance [52], and apoptosis [370] have been studied in the past.

Table 6: Sequence motifs identified in GmIPK2 protein using MyHits Motif Scan.

Motif information	No. of sites	Amino acid residues
Inositol Polyphosphate Kinase	1	16-269
Tyrosine kinase phosphorylation site	1	42-50
Protein kinase C phosphorylation site	3	35-37, 61-63, 152-154
Casein kinase II phosphorylation site	3	35-38, 100-103, 147-150
c-AMP and c-GMP dependent protein kinase phosphorylation site	1	231-234
N-myristoylation site	2	141-146, 259-264

3.3.3.3.2 Evolutionary analysis: Conserved motif analysis of IPK2 protein sequences points at a distinct evolutionary association between these kinases. The phylogenetic analysis would provide a further basis to determine their relatedness as well as to understand their collective evolution from a common ancestor. Previous research has reported that IPK superfamily of kinases to which IPK2 belongs evolved from a common ancestor [371, 372]. In our study, based on the alignment obtained in section 3.3.3.3.1 of GmIPK2 and its homologs, we constructed a neighbor-joining phylogenetic tree using MEGA 6.0 software (Figure 3.8). The tree topology derived was supported by high bootstrap values. The IPK2s were clustered into six-well delineated groups. The clusters consist of members of the Poaceae, Brassicaceae, Malvaceae, Rosaceae, Solanaceae, and Fabaceae families. The Poaceae family of monocots (*Zea may, Sorghum bicolor and Aegilops tauschii*) was found to be most distantly related to *G. max* in comparison to the Fabaceae family of eudicots (*Glycine soja, Phaseolus vulgaris, Vigna radiata, Cajanus cajan, Trifolium pratense, Cicer arietinum,*

Medicago truncatula, Lotus japonicas, *Arachis duranensis*, and *Arachis hypogaea*) which are most closely related. *Glycine soja, V. radiata, P. vulgaris,* and *C. cajan* are orthologs of *G. max* (bootstrap value 0.979; 98%, 74%, 71%, and 71% pairwise similarity respectively).

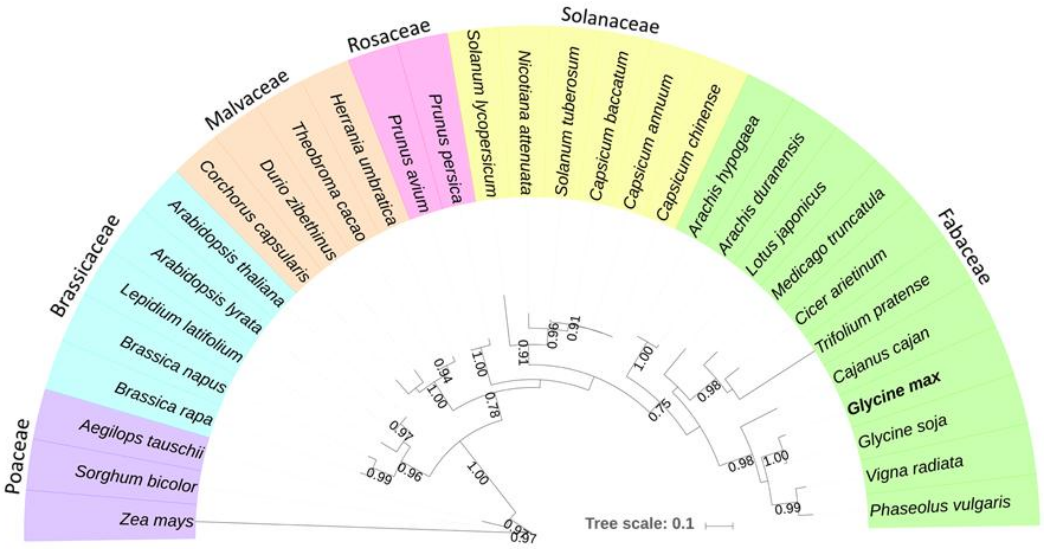


Figure 3.8: Phylogenetic tree showing the evolutionary relationship of GmIPK2 with other plant IPK2s. The sequences are divided into five clades and color coded to indicate the plant family to which they belong. The posterior probability values are indicated corresponding to every node.

3.3.4 Secondary structure prediction

The secondary structure of a protein is more conserved than its nucleotide sequence and is, therefore, a prized source of information in understanding its classification, function, molecular evolution and interaction with macromolecules [373]. In addition, the secondary structure provides the first framework for homology-based prediction of a protein 3D-model. Thus, in the current study, we inferred the secondary structure composition of IPK2 kinases from a three-state prediction done by using NPS@ web server (Table 7).

A high coil content was observed in most of the sequences including GmIPK2 while

some showed them in nearly equal proportion with α-helix. This structural state can be justified based on the rich content of highly flexible glycine and kink inducing proline amino acid residues.

Table 7: Three-state description of secondary structure content and disulfide pattern prediction of IPK2 sequences. The secondary structure data was generated by joint prediction with SOPMA and PHD while disulfide bonding pattern was determined using DiANNA (DiAminoacid Neural Network Application) 1.1 server.

Organism	α-Helix (%Hh)	Extended strands (%Ee)	Random coil (%Cc)	Disulphide bridge prediction
Glycine max	26.88	13.26	31.54	94-121, 123-276, 158-186
Glycine soja	34.27	29.78	35.96	20-117, 22-85, 161-175
Vigna radiata	29.72	17.13	24.48	127-190, 263-281
Cajanus cajan	25.84	13.11	29.21	119-177
Phaseolus vulgaris	24.62	21.21	28.79	116-176, 208-249
Lotus japonicus	25.44	14.49	26.86	94-215, 149-187, 261-283
Cicer arietinum	21.96	14.53	28.72	122-185, 217-279, 260-274
Medicago truncatula	18.54	15.02	31.69	98-262, 325-357
Arachis hypogaea	24.58	14.48	33.67	120-188, 223-288, 269-292
Arachis duranensis	24.58	14.48	33.33	223-288, 269-292
Corchorus capsularis	19.52	20.21	25.34	120-187, 222-267
Theobroma cacao	19.74	16.78	39.14	50-120, 187-222, 267-298
Herrania umbratica	18.69	14.75	42.95	120-296, 222-298, 267-288
Durio zibethinus	18.98	22.37	26.10	93-120, 149-222, 267-288
Solanum lycopersicum	18.03	18.37	30.27	120-187, 222-267
Solanum tuberosum	17.40	17.65	26.23	120-187, 222-283
Brassica rapa	13.50	21.17	22.63	144-186, 221-264
Capsicum baccatum	17.65	17.38	30.75	120-187, 222-283
Nicotiana attenuata	21.47	15.76	25.82	156-223, 258-303
Capsicum annuum	19.52	16.58	30.48	120-187, 222-267
Capsicum chinense	20.32	16.31	30.21	120-187, 222-267
Brassica napus	18.25	18.25	24.56	120-220, 144-185, 264-281
Arabidopsis lyrata	22.00	17.00	23.33	120-187, 222-272
Prunus avium	17.44	17.44	27.05	120-222, 156-187
Trifolium pratense	21.61	15.75	31.87	173-205, 203-250
Arabidopsis thaliana	22.00	16.67	24.67	120-187, 197-272
Prunus persica	17.02	17..02	29.79	120-222, 156-187
Lepidium latifolium	17.85	17.17	28.62	78-269, 120-187, 222-285
Zea mays	16.25	10.42	44.17	101-203, 192-200

Aegilops tauschii	28.57	7.67	31.36	110-188, 124-267
Sorghum bicolor	30.12	12.42	26.40	13-300, 137-160

The percentage of extended strands (%Ee) in all the kinases were found ranging from 13-29% except for the Poaceae family kinases which showed a low %Ee conformation (below 13%). PDBSum tool PROMOTIF analysis of GmIPK2 polypeptide identified total of eight α-helices and eight β-strands arranged to form three antiparallel β-sheets, interspersed throughout by regions of coil or turn conformations (Figure 3.9).

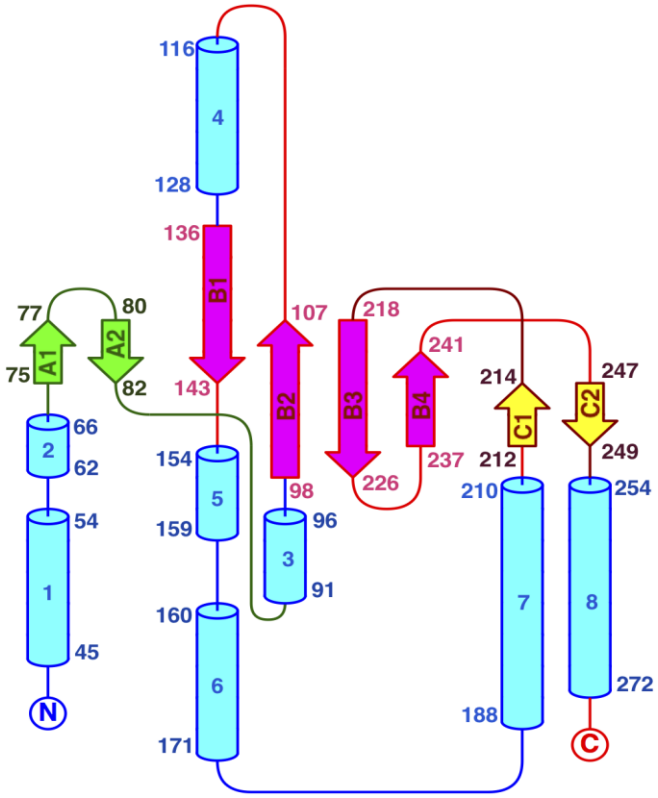


Figure 3.9: Topology map of GmIPK2 generated using ProMotif and Topdraw. There are a total of 8 α-helices (1-8) and 3 β-sheets (β-sheet 1: strands A1 and A2; β-sheet 2: strands B1, B2, B3, and B4; β-sheet 3: strands C1 and C2).

We also recognized a varying number of bonded half-cystine pairs in all the IPK2 protein sequences using DIANNA server. It revealed the presence of 8 Cys residues in GmIPK2 and the most probable half-cystine pairs predicted by CYS-REC were 94-121, 123-276 and 158-186. These potential long-term disulfide interactions participate

in stabilizing the native conformations of our proteins and may as well contribute to the differences in their tertiary structures.

3.3.5 The modeled 3D structure of GmIPK2

3.3.5.1 Model building: To derive structural information about GmIPK2 protein we built its theoretical model by following the homology modeling approach since no X-ray crystal or NMR structure of it is available. The homology modeling technique takes advantage of structural conservation found in similar proteins that have evolved from a common ancestor. Yeast IPK2 protein was the first member of the inositol multikinase family whose crystal structure was determined [362], but it was found to show very low sequence similarity with GmIPK2 protein. As we know, aligning two sequences can be difficult in case of low sequence similarity, we decided to use the GmIPK2 protein sequence as a query in PSI-BLAST to find sequences amongst the PDB database of proteins with resolved 3D structures to use as a potential template. The only closest homologous sequence available in PDB was that of chain A of *Arabidopsis thaliana* inositol phosphate multikinase (PDB: 4FRF) which showed 55% sequence identity with an e-value of 9e-99. The initial comparative models i.e. GmIPK2-S and GmIPK2-P were built using fully automated SWISS-MODEL and PHYRE2 servers respectively which also identified 4FRF_A as the most reliable template using sensitive hidden Markov model searches and used the same as the structural input. A global quality estimation score (GMQE) of 0.62 was provided by SWISS-MODEL which indicates a reasonably reliable structure. The 3D structure of GmIPK2-S was visualized by using PyMol molecular graphic system (Figure 3.10).

A homology model was additionally built using MODELLER 9.16 program from the X-ray crystal structure coordinates of the previously identified template structure (4FRF_A). The software generated five different models by optimizing the objective

function of spatial restraints in cartesian space. Three different energy scores viz. Molpdf, DOPE, and GA341 were computed for each of these models and compared to one another to select the best 3D structure (Table 8).

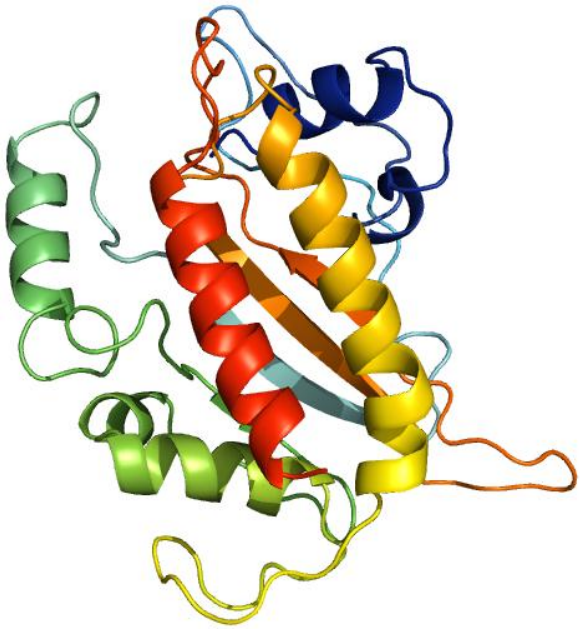


Figure 3.10: Homology model of *GmIPK2* protein generated using Swiss-model and rendered using PyMOL. The structure obtained was verified using ProSA and Verify3D servers.

Table 8: Summary of GmIPK2 models produced using Modeller 9.16.

Filename	molpdf	DOPE score	GA341 score
GmIPK2-M1	1904.96094	-26781.98828	1.00000
GmIPK2-M2	1823.05139	-27677.49023	1.00000
GmIPK2-M3	1818.68884	-26874.69922	1.00000
GmIPK2-M4	1734.48291	-27846.58203	1.00000
GmIPK2-M5	2397.21655	-26359.73828	1.00000

Model 4 (GmIPK2-M4) with the least Molpdf energy of 1734.48291 based on restraint violations and a DOPE score of -27846.58203 was chosen as the principal conformation. Some natively disordered regions with high flexibility could however not be modelled. Therefore, a sensitive prediction of disorder probability of each residue was made by PrDOS server which reported 18.28% of disordered residues at a

false positive rate of 15% (see Appendix I Figure 3). Instability was found largely confined to both the terminals of the protein suggesting its possible role in molecular recognition [373, 374].

3.3.5.2 Structure validation: We then assessed the accuracy and reliability of all the three predicted models viz. GmIPK2-S, GmIPK2-P, and GmIPK2-M4 using various online diagnostic tools. RAMPAGE server which evaluates the 3D-structures based on Ramachandran plot calculations showed variable distribution of torsion angles in all the models. Figure 3.11A shows the Ramachandran plot for GmIPK2-S model, with 92.5% residues in the favourable region, 4.9% in the allowed region and, just 2.7% residues in outlier region of the plot which reflects its superior backbone geometry.

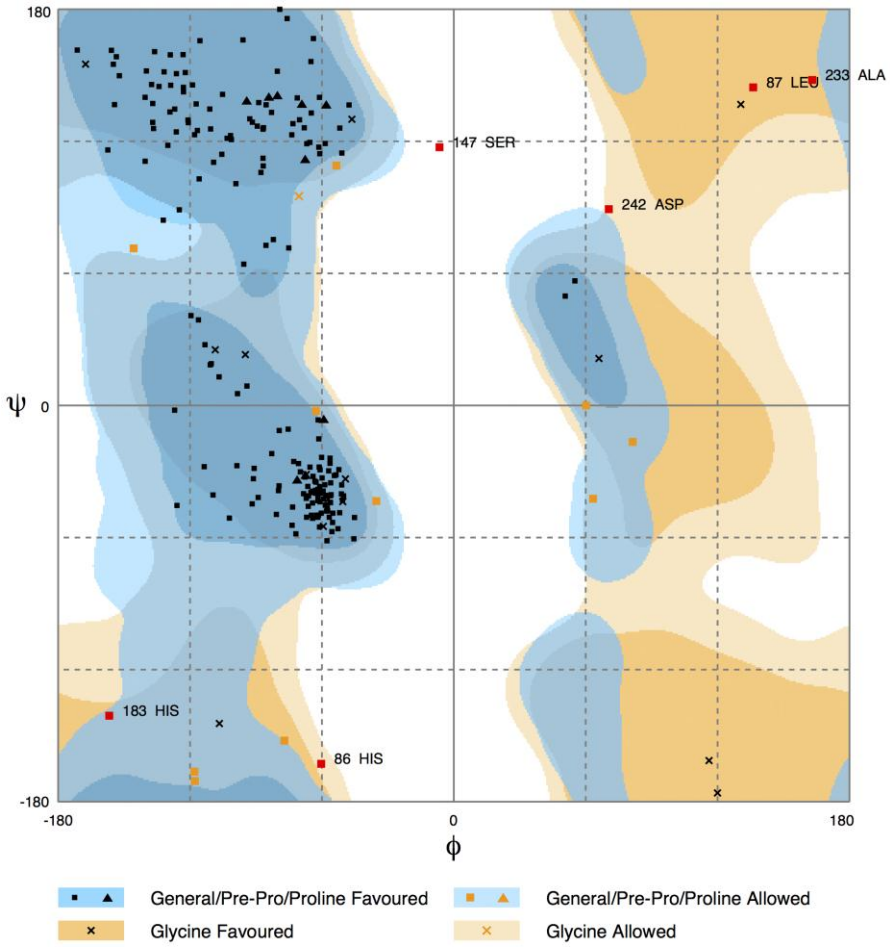


Figure 3.11: (A) Ramachandran plot generated by RAMPAGE server, validating back -bone dihedral angles of the energy minimized model.

We then utilized ProSA-web computational engine to analyze the overall quality of the models based on their z-scores and local quality based on their residue energies. GmIPK2-S model showed a better z-score of -6.56 as displayed in the energy distribution plot derived from other experimentally determined protein structures of similar size which indicates it is very much within the range of native conformations of this group (Figure 3.11B). Moreover, its residue energy computed with a window size of 40 was found to be highly negative (Figure 3.11C). We further analyzed the accuracy of 3D-models from their energy profiles obtained by Verify3D program. Figure 3.11D shows 3D-1D profile for GmIPK2-S model with 97.82% of the model residues showing an average 3D-1D profile score >= 0.2 and hence validate that majority of its amino acid sequence reconcile to its environment in the 3D structure.

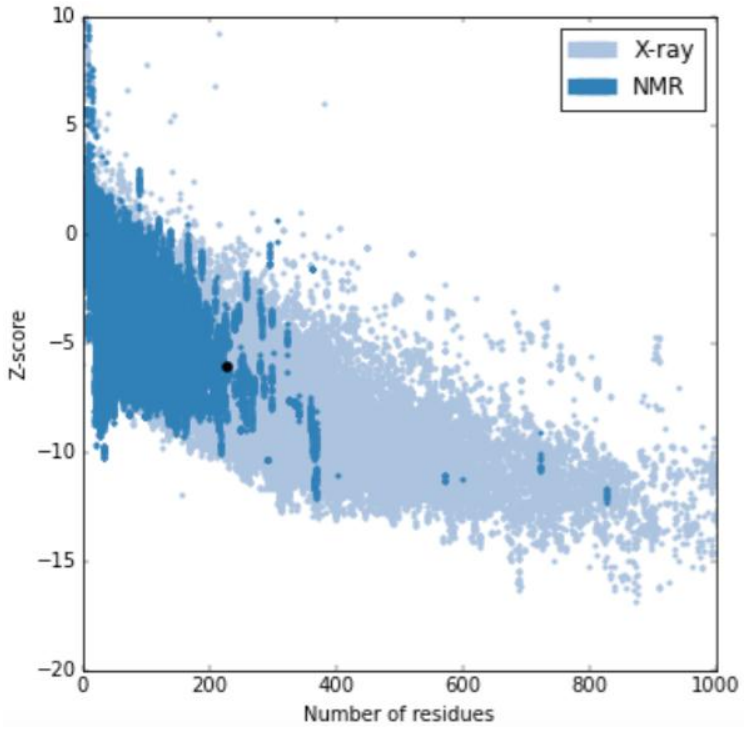


Figure 3.11: (B) Z-score plot generated using ProSA program showing a z-score of -6.56 indicating the overall quality is within the range of scores typically found for native proteins of similar size.

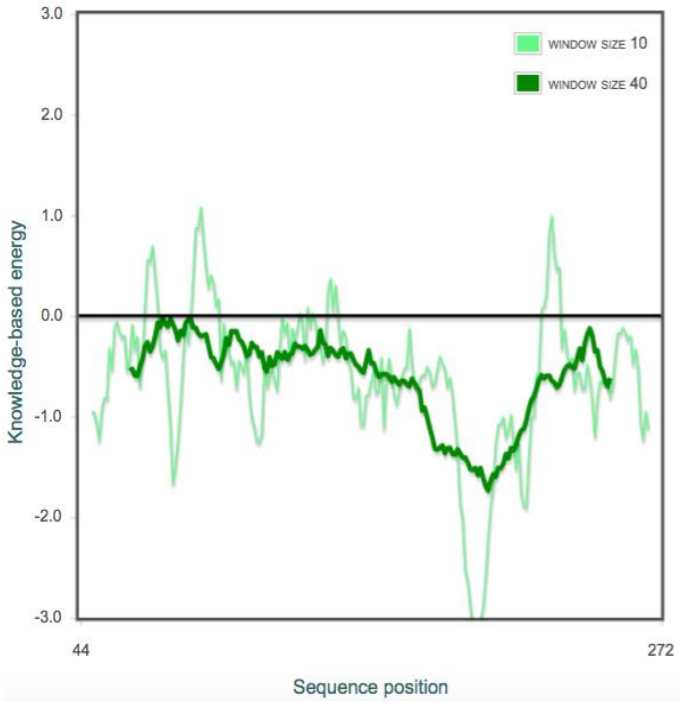


Figure 3.11: (C) Energy plot also generated using the ProSA program showing negative energy values throughout the sequence indicating a good local model quality.

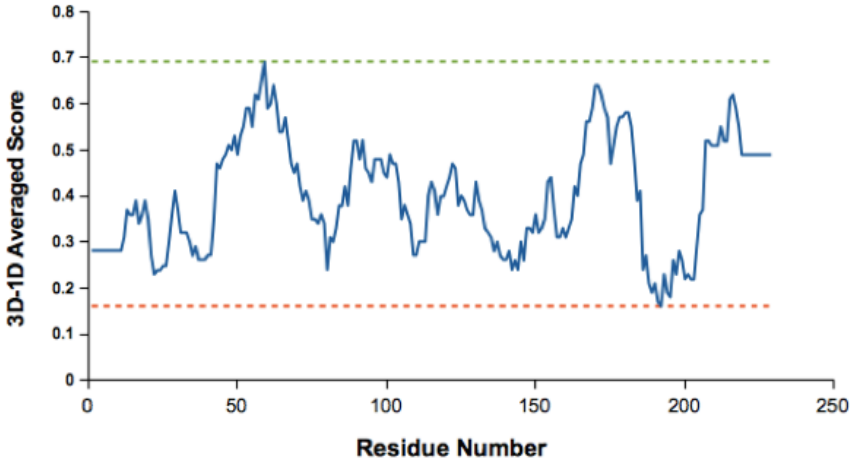


Figure 3.11: (D) 3D-1D profile showing the compatibility of GmIPK2-S model structure with its amino acid sequence. Residues showing positive compatibility scores are reasonably folded.

Additionally, the pairwise 3D structural alignment of GmIPK2-S model with the

template protein (4FRF_A) done using MATRAS 2.1 program revealed that both the structures shared 91.3% secondary structure identity (Figure 3.12A) and the average distance between the Cα backbone atoms of their 3D structures i.e. root mean square deviation (RMSD) measured through superimposition was 0.44 Å (between 196 atom pairs) (Figure 3.12B). Based on the majority of winning scores, GmIPK2-S was chosen as the best comparative model for energy minimization and further analyses.

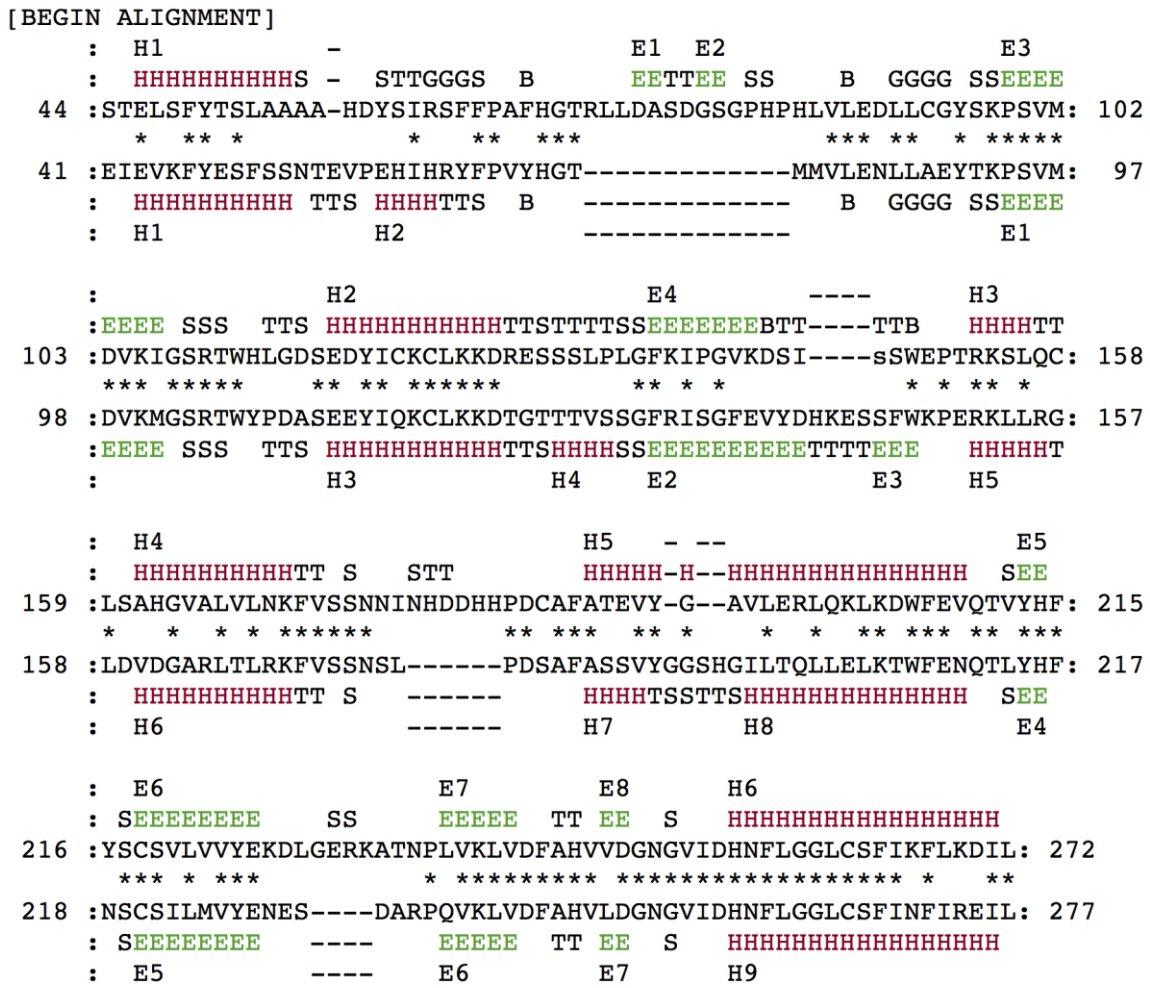


Figure 3.12: (A) 3D structural alignment between model (GmIPK2-S) and the template protein (4FRF_A) generated using MATRAS server showing 91.3% secondary structure identity.

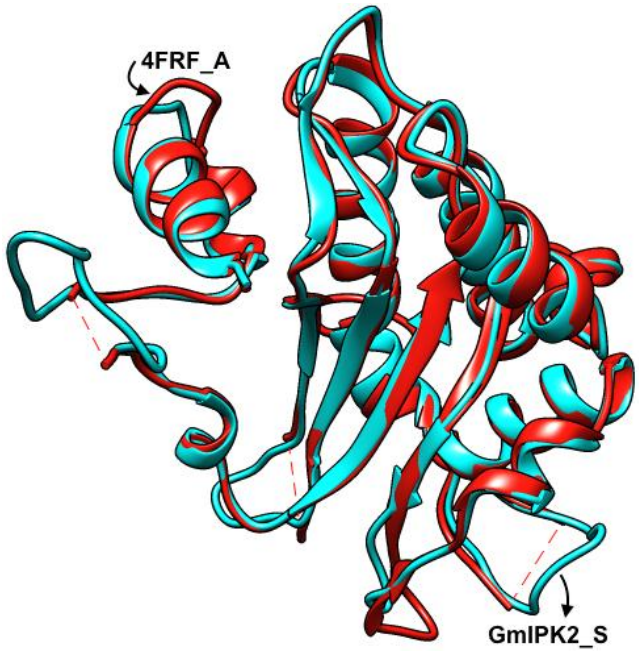


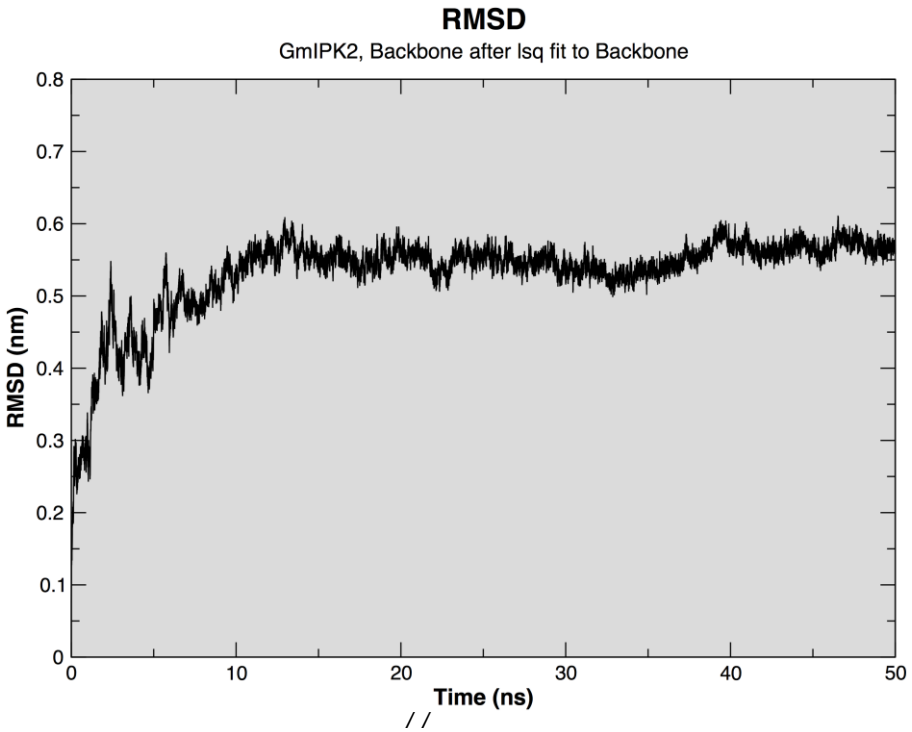
Figure 3.12: (B) Superimposed 3D structures of template 4FRF_A protein (red) and refined GmIPK2_S model (blue).

3.3.5.3 Molecular dynamics simulation

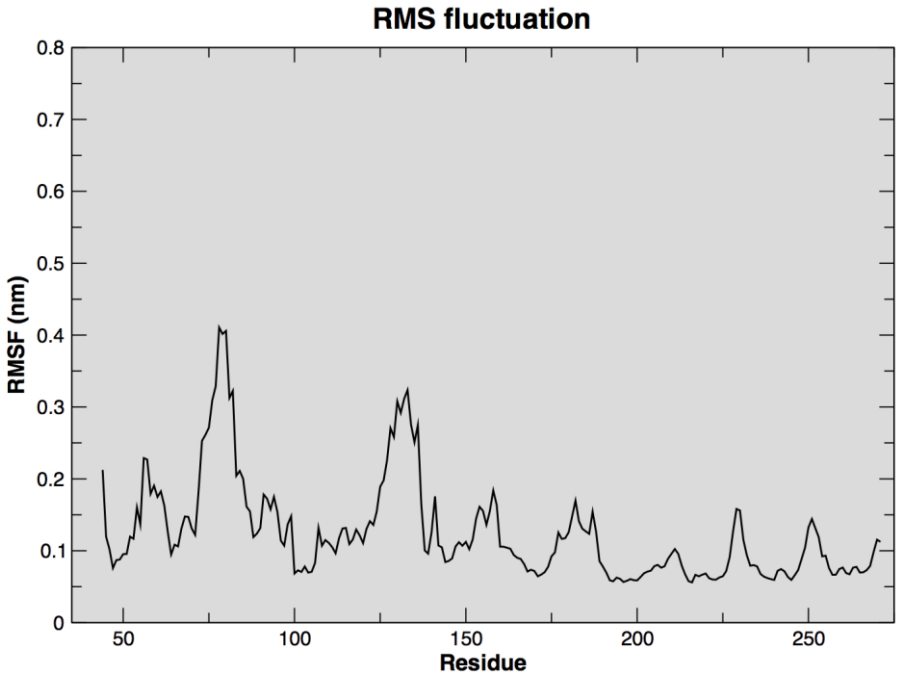
3.3.5.3.1 Trajectory files generation: We subsequently performed MD simulation on our predicted GmIPK2-S model using the GROMOS96 53A6 force field to compute its stability and dynamics. Initial potential energy minimization of the solvated model showed that the maximum force dropped below the defined value of 1000 kJ mol^{-1} nm^{-1} in nearly 500 steps. The protein structure was then subjected to 50 ns simulation run at a constant temperature and pressure to obtain the molecular trajectories.

3.3.5.3.2 Root mean square deviation and Root mean square fluctuation plot analysis: These molecular trajectories were used to compute RMSD of Cα backbone atoms of the model using the starting structure of simulation as a reference. Figure 3.13A shows RMSD for each run as a function of simulation time.

No significant change was observed in the pattern of deviation corresponding to each simulation run. There were no detectable intramolecular clashes after the minimization process. We observed that the protein stabilized around 10 ns of the production run and converged to ~0.57 nm at 50 ns. The initial RMSD increase can be attributed to the restraints applied to the system during the equilibration phase and, their release later at the beginning of the production phase. Besides RMSD, we also calculated root mean square fluctuations (RMSFs) to monitor motility of residues over the simulation time to draw an idea of its flexibility regions. From the RMSF plot (Figure 3.13B), we identified Leu (75, 135, 272), Asp (76, 79, 127), Ala (77), Ser (78, 130-132), Gly (80, 136), His (84), Lys (126, 154), Glu (129, 230), and Arg (128, 153) residues showing higher movement from their native position. In other words, these residues are dynamic in nature and therefore functionally more relevant. We also converted these RMS fluctuations to B-factor values to highlight residues with dynamic mobility in the final 3D model of the protein (Figure 3.13C). Thus, overall, simulation results highlight the stable nature of our protein model and find it reliable to be used for further active site predictions.



(A)



(B)

Figure 3.13: (A) Root mean square deviations and **(B)** Root mean square fluctuations of the Cα backbone atoms in GmIPK2_S model over 50 ns MD simulation.

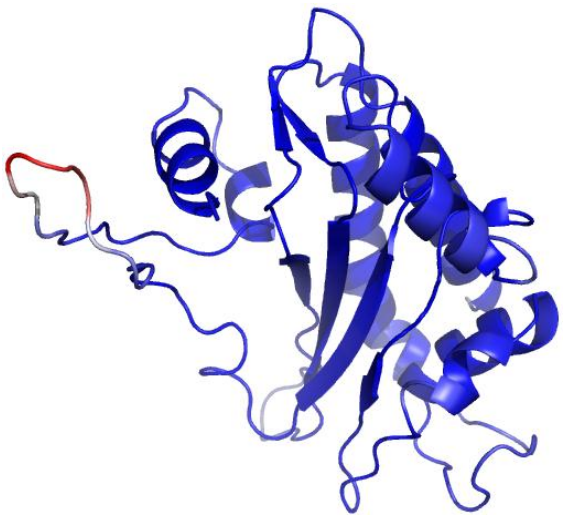


Figure 3.13: (C) Protein backbone coloured in a blue-red-white gradient with the B-factor values indicating most flexible regions of the protein in red.

3.3.6 Molecular docking with inositol phosphates

3.3.6.1 Active site prediction: Once the protein model was refined, a comparative study was performed to detect possible binding pocket residues using CASTp, FTSite and FunFOLD2 binding site prediction servers (see Appendix I Figure 4). The predictions made by these servers helped in reducing the amount of sampling done during the docking procedure. Lys 105, Thr 110, Lys 122, Lys 126, Ser 130, Lys 138, Ile 139, Pro 140, Arg 153, Lys 154, Gln 157 and Ser 219 were determined as possible active site residues in GmIPK2-S model. Similar predictions were also made in template protein (4FRF_A) which identified Arg 104, Thr 105, Pro 108, Phe 137, Lys 149, Arg 152, His 216, Asn 218, Ser 219, Gln 242 and Val 246 as the probable binding site residues. From these studies, we deciphered that Lys, Arg, Pro, Thr, Gln, and Ser residues are highly conserved in active sites of functionally identical model and template proteins.

3.3.6.2 Docking and residue interaction analysis: Molecular recognition is vital to many biological processes. However, experimental determination of structures of molecular interactions is cost intensive, demand time and expertise. We, therefore, chose computational molecular docking to model our protein-ligand binding and characterize the interactions between its binding pocket residues and known active ligands. In addition, we further evaluated the quality of the constructed homology model with the aid of molecular docking by checking whether the model contains protein-ligand contact suggested by experiments. Multiple substrate specificities have been described previously for IPK2 gene product, primarily catalyzing 5GUG-I3P and 4A69-I0P [108]. We thus docked these centroid ligands into the binding cavity of GmIPK2 protein using molecular docking program Autodock Vina based on a semi-flexible docking approach with the scaling factor defined within 0.1 nm to predict

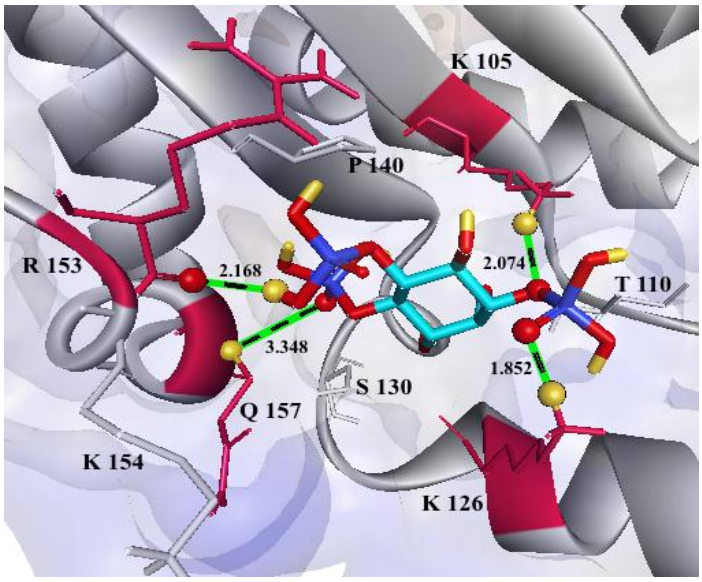
their bound geometry. VINA uses its iterated local search global optimizer algorithm to produce 9 different poses of which pose 1 corresponding to each ligand was identified as the best binding mode based on their lowest binding affinity score of -6.2 kcal/mol for 5GUG-I3P and -5.8 kcal/mol for 4A69-I0P computed by VINA's default statistical scoring function (Table 9).

Table 9: Ligands with their binding affinity with the best mode highlighted in red.

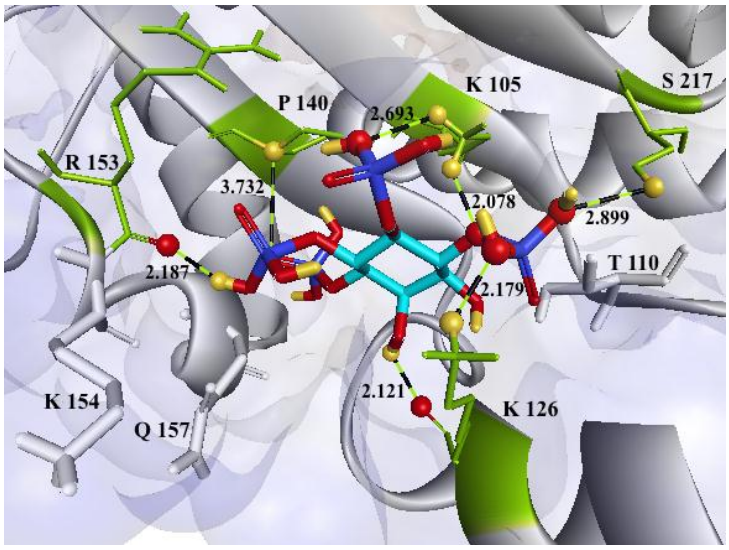
Modes	Binding affinity (kcal/mol)	
	Ligands	
	5GUG-I3P	4A69-I0P
1	-6.2	-5.8
2	-5.8	-5.2
3	-5.6	-4.9
4	-5.5	-4.9
5	-5.5	-4.6
6	-5.3	-4.5
7	-5.3	-4.5
8	-5.2	-4.5
9	-5.1	-4.3

We further inspected the molecular interactions between these protein-substrate poses using Discovery Studio to predict functionally important amino acid residues and found that both the ligands were stabilized in their active site area by strong hydrogen bonding interactions (Figure 3.14A and 3.14B). Lys 105 and Lys 126 were identified as H-donors to the phosphate group oxygen of both 5GUG-I3P and 4A69-I0P while Arg 153 an H-acceptor for hydrogen bonds formation suggesting that these binding pocket residues may play a pivotal role in enzymes function and protein structure stability. Based on the previous work conducted by Holmes and Jogl [362], we hypothesize that side chains of these amino acid residues may assist in inducing conformational changes on inositol phosphate binding, enabling the enzyme to interact with differently phosphorylated inositol polyphosphates in different orientations, thus endorsing its substrate versatility. Besides, Ser 217 also form a

stabilizing hydrogen bond with the 4A69-I0P ligand. Gln 157 and Pro 140 were observed to form an unconventional carbon-oxygen hydrogen bond with 5GUG-I3P and 4A69-I0P respectively indicating their possible contribution to ligand binding affinity and ligand recognition [376].



(A)



(B)

Figure 3.14: Molecular docking of substrates to the GmIPK2_S model. Hydrogen bonding interactions of **(A)** 5GUG-I3P **(B)** 4A69-I0P substrates with residues in the active site of the protein. The substrate is depicted in stick representation with carbon atoms coloured turquoise, oxygen atoms red, hydrogen atoms yellow and phosphorus

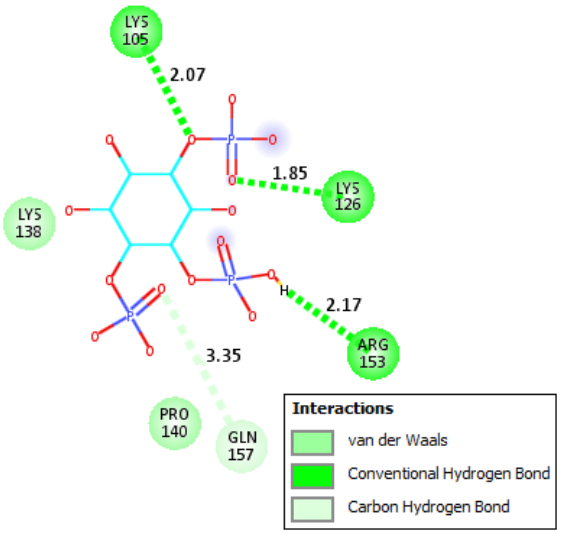
atoms blue. The interacting protein side chains are represented as maroon sticks in A and green sticks in B.

Table 10A: Hydrogen bonds between the active site residues of GmIPK2 and its substrate 5GUG-I3P along with their distances and angles.

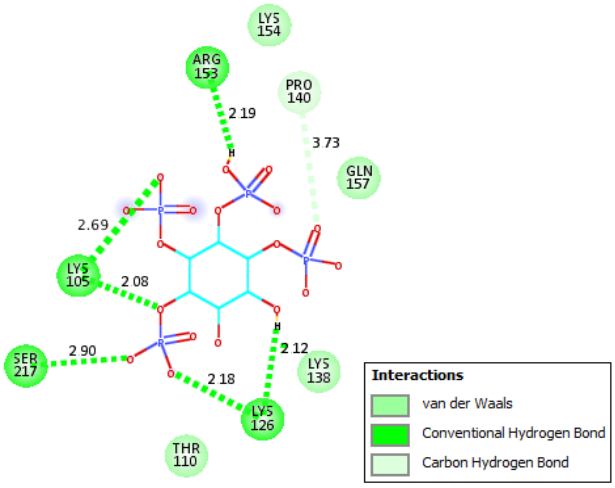
GmIPK2			5GUG-I3P		Distance	Angle
Residue	**Atom**	**Chemistry**	**Atom**	**Chemistry**	**(A°)**	**(Degree°)**
Lys105	HZ3	H-Donor	O7	H-Acceptor	2.074	49.9029
Lys126	HZ3	H-Donor	O14	H-Acceptor	1.852	44.2482
Arg153	O	H-Acceptor	H39	H-Donor	2.168	45.6080
Gln157	CA	H-Donor	O18	H-Acceptor	3.348	52.1869

Table 10B: Hydrogen bonds between the active site residues of GmIPK2 and its substrate 4A69-I0P along with their distances and angles.

GmIPK2			4A69-I0P		Distance	Angle
Residue	**Atom**	**Chemistry**	**Atom**	**Chemistry**	**(A°)**	**(Degree°)**
Lys105	HZ2	H-Donor	O18	H-Acceptor	2.693	47.8602
Lys105	HZ3	H-Donor	O24	H-Acceptor	2.078	28.0171
Lys126	HZ3	H-Donor	O28	H-Acceptor	2.179	35.7367
Lys126	O	H-Acceptor	H40	H-Donor	2.121	39.3959
Pro140	CA	H-Donor	O4	H-Acceptor	3.732	60.3277
Arg153	O	H-Acceptor	H36	H-Donor	2.187	16.0359
Ser217	HG	H-Donor	O27	H-Acceptor	2.899	22.5987



(C)



(D)

Figure 3.14: (C) & **(D)** 2D-schematic representation of the interactions shown in Figure 3.14 A & B respectively drawn using Discovery Studio Visualizer.

Lengths and angles of hydrogen bonds stabilizing the GmIPK2-I3P and GmIPK2-I0P complexes are enlisted in Table 10A and Table 10B respectively. Moreover, our analysis indicates that Thr 110, Lys 138 and Lys 154 show non-bonding interactions with the ligands. 2D protein GmIPK2-I3P/I0P ligand interaction diagrams (Figure 3.14C and 3.14D) were also generated using Discovery Studio.

3.3.6.3 Molecular dynamics simulation of GmIPK2-I3P and GmIPK2-I0P complexes: We subsequently subjected our docked protein-ligand complexes to MD simulation using the GROMOS96 43A1 force field to compute their stability and dynamics. The trajectories obtained were utilized to construct their respective RMSD and RMSF plots. The GmIPK2-I3P and GmIPK2-I0P complexes exhibited a deviation between ~0.5 to 0.6 nm (Figure 3.15A and Figure 3.16A) that converged to ~0.58 nm and ~0.55 nm respectively at 50 ns. This suggests stabilization of the structures post-simulation.

(A)

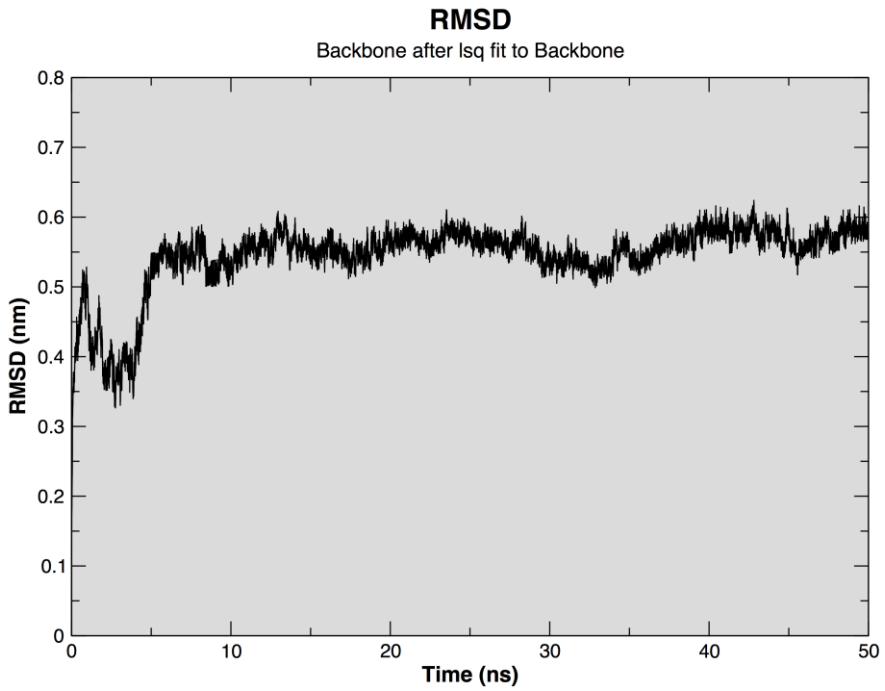
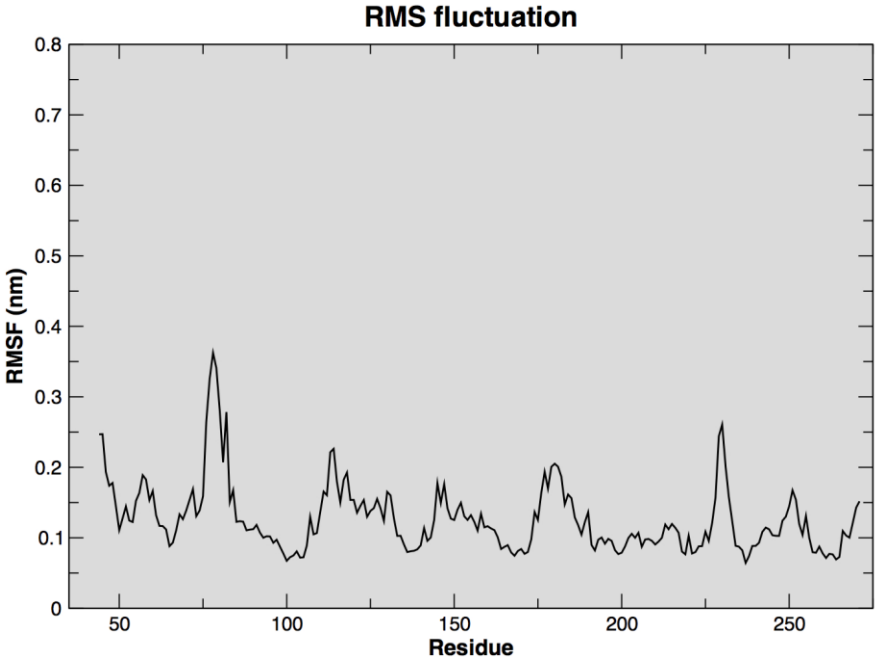
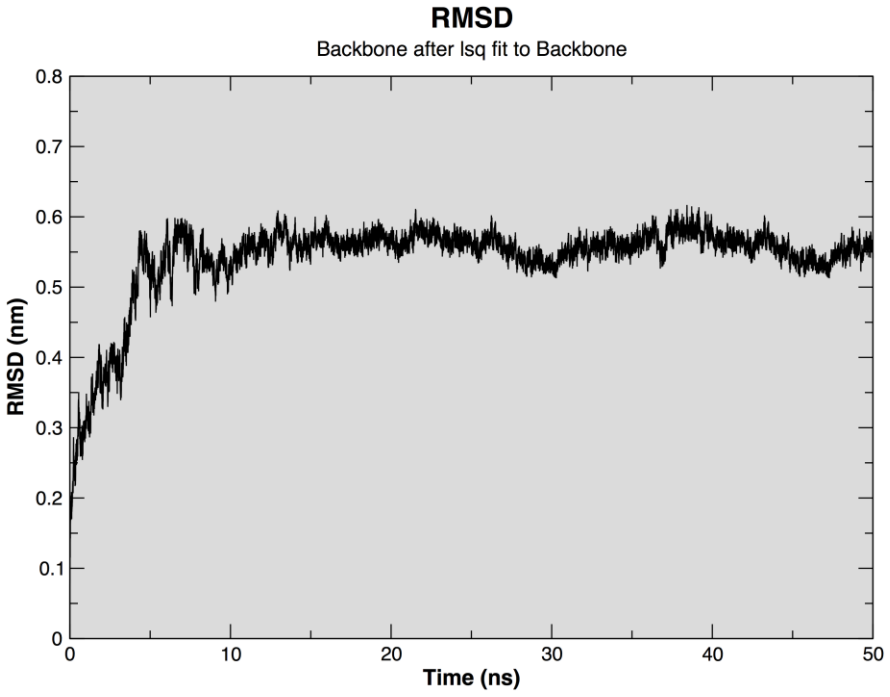


Figure 3.15: Simulation behavior of GmIPK2 complexed with 5GUG-I3P over 50 ns MD simulation. **(A)** Root mean square deviations.



(B)

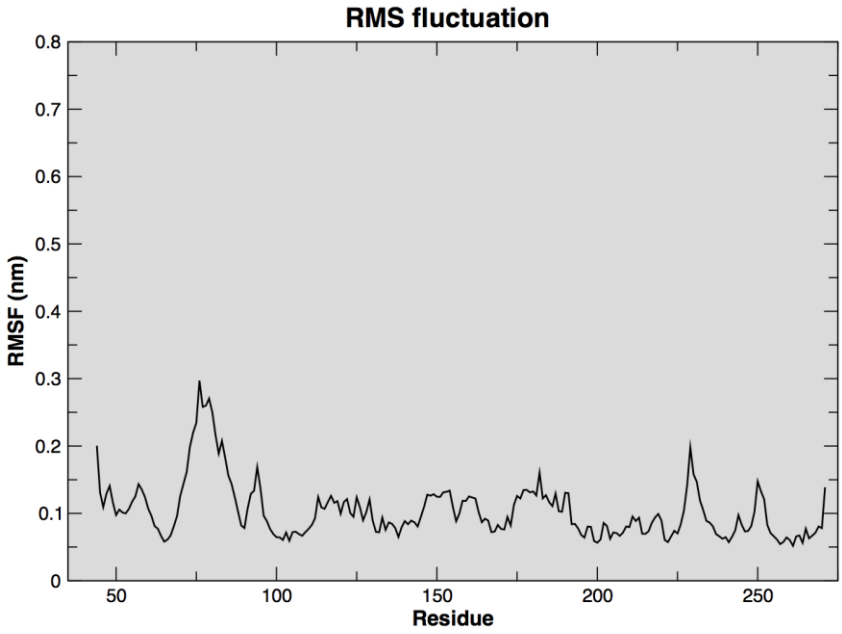
Figure 3.15: Simulation behavior of GmIPK2 complexed with 5GUG-I3P over 50 ns MD simulation. **(B)** Root mean square fluctuations.



(A)

Figure 3.16: Simulation behavior of GmIPK2 complexed with 4A69-I0P over 50 ns

MD simulation. **(A)** Root mean square deviations.



(B)

Figure 3.16: Simulation behavior of GmIPK2 complexed with 4A69-I0P over 50 ns MD simulation. **(B)** Root mean square fluctuations.

From their RMSF plots, we observed fluctuations up to ~0.38 nm in the GmIPK2-I3P complex (Figure 3.15B) and ~0.29 nm in the GmIPK2-I0P complex (Figure 3.16B) which reveal the characteristic regional flexibilities of functional significance in each complex. Lastly, we calculated the binding free energies ($\Delta G_{binding}$) of both the complexes by using *g_mmpbsa* for their respective trajectories. $\Delta G_{binding}$ values of -12.4 kcal/mol and -11.9 kcal/mol were estimated for GmIPK2-I3P and GmIPK2-I0P complexes respectively. These values indicate that the binding of ligands is thermodynamically favourable and thus validates the reliability of our simulation.

3.4 Conclusion

The present work elucidates molecular characteristics, structure, and function of *GmIPK2* gene and the protein. To conduct this study, we isolated and cloned a partial sequence of the gene from *G. max* cv. Pusa16 and analyzed it using computational tools. Structural annotation of the gene highlighted seed-specific cis-acting regulatory motifs in its promoter region. Further assessment of its transcripts verified its predominant expression in the seeds, specifically during later stages of development. We thus speculated that perturbing *GmIPK2* gene expression in seed tissue could be a viable strategy for reducing PA biosynthesis. Sequence similarity search of the translated GmIPK2 ORF identified homologous proteins with shared functions. Alignment of these homologs located small conserved motifs which were further confirmed by the evolutionary analysis. Motif analysis characterized GmIPK2 as a tyrosine-specific protein kinase which participates in different cellular processes such as protein-protein interactions, cell signaling regulation, cell adhesion, and transcription. These motifs also predicted cytoplasmic localization of the GmIPK2 protein. Phylogenetic tree recognized IPK2 protein from *G. soja*, *V. radiata, P. vulgaris,* and *C. cajan* as the closest homologs of GmIPK2 which indicate that they are functionally very similar. Since protein structure is the foundation for understanding its function, we determined the 2D and 3D homology model of GmIPK2 protein and performed its docking. The will assist in making structure-based predictions of its function such that based on the 3D structure of an active site motif as well as in determining its empirical model. The docking studies will also help in designing potent IPK2 inhibitors for studying PA biosynthesis pathway in detail. In summary, these results provide preliminary data for achieving the main objective of reducing PA content in soybean seeds to enhance its nutritional quality, addressed in the following section.

This work has been published in "*3 Biotech*", Springer, entitled "***Molecular characterization, modeling, and docking analysis of late phytic acid biosynthesis pathway gene, inositol polyphosphate 6-/3-/5-kinase, a potential candidate for developing low phytate crops***".

Chapter 4

Development and evaluation of low phytate (lpa) soybean by targeting seed-specific silencing of *inositol polyphosphate 6-/3-/5-kinase* gene via encoding a self-complementary hairpin RNA

4.1 Introduction

Soybean represents one of the most important, inexpensive sources of energy-dense and nutrient-rich vegetarian food supply to the world population. In many countries, soy meal is a major source of protein and minerals for animals. Its high nutritional value comes from the accumulation of many beneficial storage compounds during its seed development. However, along with these compounds, there is a corresponding accumulation of molecules with anti-nutritional properties. Phytic acid amongst them is one of the most prominent antinutrients found as it represents the most abundant form of phosphorus occurring in its seeds (up to 85%) and other tissues and organs. Owing to its chemical structure, PA function as a chelator of nutritionally important positively charged macro and micro minerals and form stable salts with them called phytins [5, 377]. Due to lack of phytase activity in the gut of monogastric animals, most of the phosphorus and minerals in these salts remain unutilized and are excreted, potentially contributing to nutritional deficiency in human and animal populations [6, 98, 378, 379]. In addition, animal waste rich in phytins is used as a ready source of fertilizer, causing accumulation of excess phosphorus in the soil which contributes significantly to eutrophication of surface waters [139, 380]. To overcome the nutritional deficiencies foods and feeds were supplemented with available forms of dietary phosphorus. However, this doesn't solve the problem of PA derived

phosphorus pollution. Therefore, scientists began looking into approaches which obviate these problems simultaneously. Amongst them, generation of low phytate crops (lpa) was the highlight.

The first generation of lpa crops was developed using classical breeding/ mutagenesis which blocked the synthesis or accumulation of PA. These efforts were augmented with subsequent advancements in technology by the aid of forward genetics. Several lpa genotypes have been isolated for a number of crop species including maize [140, 274], wheat [158], barley [144, 381], rice [150, 151], and Arabidopsis [87]. Lpa mutant lines CX1834 [140] and LR33 [100] were developed in soybean. Some animal and human nutrition studies have shown that subjects consuming low-phytate feeds show increased phosphorus utilization and reduced phosphorus elimination in their waste as well as the increase in percentage absorption of Fe, Zn and Ca [382-385]. Unfortunately, these lpa crops thus generated were largely unsuccessful due to their poor agronomic performance which called in for some alternate approach which maintains plant quality.

Research in the past has provided evidence that manipulating plants by genetic engineering techniques hold great potential for crop improvement, and hence this approach was explored by research teams working on generating lpa crops. The first strategy that was followed to develop lpa soybean was the transgenic accumulation of microbial phytases in their seeds [16, 131]. However, the cost and the labour incurred for its processing made it an economically unviable alternative. Successively, a sustainable approach directing expression of soybean phytase during embryo development at the site of PA synthesis/storage was followed [157, 386]. Nonetheless only a maximum reduction of 25% in PA level could be attained by this method which is significantly less than what was achieved previously. Therefore, to

reduce the PA level even further albeit optimally, we explored a reverse genetics technique of RNA interference (RNAi) in our study, which was recently successfully used to generate lpa rice [387, 388] and wheat [389]. Nunes and co-workers back in 2006 [23] achieved up to 95% reduction in soybean seed PA by RNAi but at the cost of embryo abortion in the progeny zygotes. The observed effect can partly be accounted for by firstly, the choice of gene and secondly, the promoter used for expression of the transgene. To manipulate PA biosynthesis, Nunes followed inhibition of the first step of the pathway which is suggested to be the most effective strategy. However, suppressing myo-inositol-1-phosphate synthase (*GmMIPS*) expression may lead to critical alterations in seed PA biosynthesis [24], subsequently disturbing the cellular phosphorus and inositol homeostasis. Therefore, in our study we decided to choose one of the late PA pathway enzymes, inositol polyphosphate 6-/3-/5-kinase (*GmIPK2*) as the target gene and designed an intron-containing self-complementary hairpin (ihp) RNA specific to it. Due to its multiple specificities resulting in sequential phosphorylation of 1D-myo-inositol-1,4,5-trisphosphate to 1D-myo-inositol-1,3,4,5,6-penta*kis*phosphate and a strategic position in the pathway, we hypothesize that this enzyme will achieve a higher but optimal level of PA reduction. Recently, much research on the constitutive expression of ihp RNA has been reported to silence target genes efficiently in a variety of species [390]. The ability of constitutive promoters to direct high levels of transgene expression can, however, be a limiting factor when temporal and spatial gene expression control is required to achieve manipulation in specific plant organs or developmental stages [391]. Its been observed to interfere with plant development resulting in abnormal phenotypes [392, 393]. Nunes also used a constitutive promoter CaMV 35S to drive the expression of the target gene. However, he obser-

ved that it resulted in a strong silencing in vegetative tissues other than the developing seeds. From this, we concluded that the promoter we use for expression should be active only in the target tissue i.e. developing seeds, the storage site of PA which will help in minimizing any off-target effects. Also, we kept in mind that the promoter we use should have the same temporal and spatial activity in the seed as our target gene *GmIPK2* to achieve its successful silencing. To solve our purpose, we selected the promoter of reserve protein vicilin, located in protein bodies which is the same site as *GmIPK2* gene expression to drive our RNAi construct.

In summary, in the present study, we attempted the use of RNAi based ihp RNA to silence *GmIPK2* gene expression in soybean seeds using a seed-specific promoter vicilin and achieve a reduction in its seed phytate levels and a corres-ponding increase in available phosphorus and some PA bound minerals.

4.2 Materials and Methods

4.2.1 Seed material

Glycine max [L.] Merr. cv. Pusa-16 was procured from Division of Genetics, IARI, New Delhi, India for isolating *GmIPK2* gene as described in the previous chapter and as the recipient genotype for subsequent genetic modifications. Quantitative character -istics of this cultivar grown primarily in northern plain and hill zones of India are described by Karnwal and Singh [394] and Ramteke and co-workers [395, 396].

4.2.2 Vectors

IPK2 sense, antisense, and intron fragments were generated initially in pCR^{TM}2.1-

TOPO® TA cloning vector. The final ihp construct was sub-cloned in the binary vector pCWAK (**Figure 4.1**) prepared in-house in a previous work by ligating fragments from the expression vector pCW66 containing seed-specific promoter vicilin and binary vector pAKVS with plant selectable marker *bar* gene. Both the vectors were obtained from Dr. Craig Atkins Laboratory, Department of Botany, School of Plant Biology, University of Western Australia, Australia.

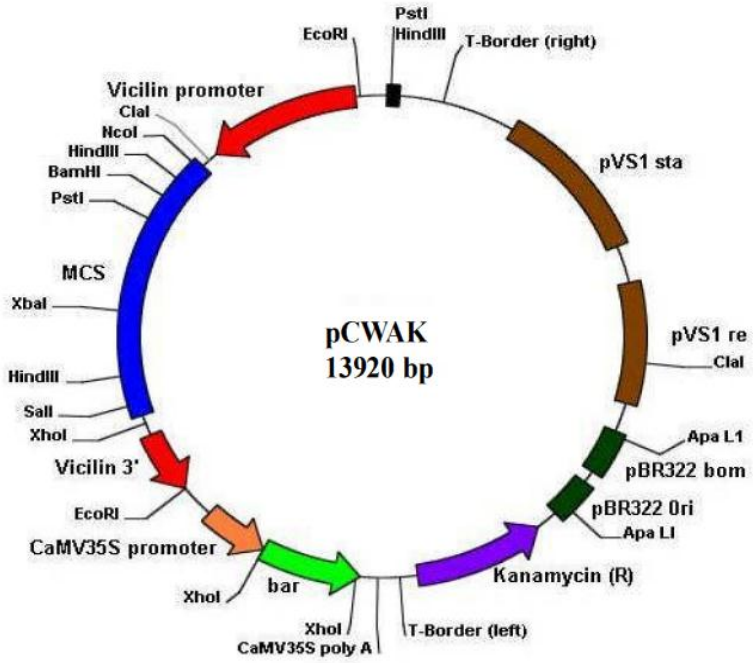


Figure 4.1: Circular map of binary vector pCWAK containing bacterial and plant selectable marker genes kanamycin and *bar* respectively as well as seed-specific promoter vicilin.

4.2.3 Preparation of *GmIPK2* specific ihp silencing construct

When designing an ihp RNAi construct, one of the biggest challenges is to an obtain inverted versus direct orientation of the double-stranded target RNA fragment. In order to overcome this challenge, we adapted directional cloning of the sense, antisense, and intron cDNA fragments by incorporating pairs of compatible restriction enzymes in their respective PCR amplicons. Restriction of these enzymes generates compatible cohesive

ends that increase the frequency of obtaining correct inverted repeats.

4.2.3.1 Cloning of *GmIPK2* sense sequence

For cloning *GmIPK2* in sense orientation (*GmIPK2_S*), a 305 bp cDNA fragment (including 250 bp of its 3' end coding sequence and a 55 bp conserved sequence from its 3' untranslated region) was PCR amplified from *GmIPK2* transcript cloned previously (section 3.3.1) using gene-specific primers containing *Bam*HI and *Sac*II restriction sites in forward and reverse primers respectively.

Primers used for amplification:

GmIPK2_S FP: -5′ CGCGGATCCGCGTTGCAGAAGCTCAAG 3′

GmIPK2_S RP: -5′ TCCCCGCGGGGAGCGACACTAATTCAAG 3′

PCR reaction mixture

Reagent	Volume (µl)
10X PCR buffer	2.5
dNTP mix (10mM each)	1
25mM $MgCl_2$	2
GmIPK2_S FP (200 pmol/µl)	0.5
GmIPK2_S RP (200 pmol/µl)	0.5
Template DNA (50ng/µl)	2
TaqDNA polymerase (5U/µl)	0.5
NFW	to 25

After preparing the PCR reaction mixture, amplification was performed by placing the reactions on a PTC-100^{TM} Peltier thermal cycler (MJ Research).

PCR cycling protocol

Steps	Temperature	Time	Cycles
Initial denaturation	94^{o}C	4 min	1
Denaturation	94^{o}C	30 sec	
Annealing	65.3^{o}C	30 sec	35
Extension	72^{o}C	30 sec	
Final extension	72^{o}C	10 min	1

On completion, 2µl of the amplification product was evaluated by agarose gel electrophoresis (see Appendix II) alongside a 1 kb DNA ladder to determine the approximate size of the amplicon. The PCR reaction was gel-purified to remove undesired amplification products and primer dimers by using gel/PCR DNA Fragments Extraction Kit (DF100) from Geneaid and subsequently improve the ligation efficiency. The desired amplicon was then ligated in the TA cloning vector pGEM-T Easy (Promega, USA).

Cloning reaction

Reagent	Volume (µl)
Salt Solution	1
TOPO® vector	1
GmIPK2_S insert	1
NFW	to 6

The ligation mixture was incubated at 4^{o}C overnight and used directly for the transformation of competent *E. coli* DH5α cells (see Appendix II). The recombinant plasmids were identified by blue-white screening, isolated (see Appendix II), and confirmed by restriction digestion with *Eco*RI to release the amplicon fragment as well as by sequence analysis. The clone was named as pMS.

4.2.3.2 Cloning of *GmIPK2* antisense sequence

For cloning *GmIPK2* in antisense orientation, a 305 bp cDNA fragment (including 250 bp of its 3' end coding sequence and a 55 bp conserved sequence from its 3' untranslated region) was PCR amplified from *GmIPK2* transcript cloned previously (section 3.3.1) using gene-specific primers (GmIPK2_As; FP 5'-CCGCTCGAGCGG AACGTCTTCGAGTTC-3' and RP 5'CCATCGATGGCGCTGTGATTAAGTTCGTA-3') containing *Xho*I and *Cla*I restriction sites respectively. The PCR reaction mix and

cycling conditions are the same as that for *GmIPK2_S* fragment amplification. The PCR product was then cloned in pCR^{TM}2.1-TOPO® vector similarly as that described for GmIPK2_S. The positive clones were confirmed by restriction with *Eco*RI as well as by sequence analysis. The clone was named as pMAS.

4.2.3.3 Cloning of *GmFad2-1* intron sequence

A 395 bp soybean fatty acid desaturase genes' (*GmFad2-1*) intronic sequence containing six and nine base pair exonic sequence at its 5' & 3' ends respectively was PCR amplified using primer pairs (*GmFad2-1*; FP 5'-TCCCCGCGGGGAAGG TCTGTCTTATTTTGAATC-3' and RP 5'-CCATCGATGGTATACCGCACTAGTAA ACCAC-3') containing *Sac*II and *Cla*I restriction sites respectively and cloned in pCR^{TM}2.1-TOPO® vector in similar manner as that described for GmIPK2_S and GmIPK2_As. The clone was named as pMINT. It was used as the base vector for assembling the final *GmIPK2* ihp silencing construct.

4.2.3.4 Assembly of *GmIPK2* ihp silencing construct

GmIPK2 ihp silencing construct is put together in two stages: 1) Ligation of gene-specific sense and intron fragments, 2) Ligation of antisense fragment to sense-intron ligated fragments to form the complete silencing cassette.

4.2.3.4.1 Ligation of GmIPK2_S and GmFad2-1 fragments

Both pMS and pMINT vectors were digested independently with *Bam*HI and *Sac*II restriction enzymes in double digestion to facilitate directional cloning of GmIPK2_S and GmFad2-1 fragments. The base vector pMINT was first digested with the enzyme

pair to linearize the plasmid with *Bam*HI and *Sac*II overhangs. The pMS vector was then digested with the same enzyme pair to generate the same overhangs on the cloned GmIPK2_S PCR fragment. The GmIPK2_S with compatible cohesive ends was subsequently ligated into the linearized pMINT vector. The restriction reaction common to both the digestions were prepared as follows:

Restriction mixture

Reagent	Volume (µl)
pMINT/pMS (1-2 µg)	5
10X NEBuffer 4	2
*Bam*HI (20U/µl)	1
*Sac*II(10 U/µl)	2
NFW	to 20

The reaction mixtures were incubated at 37^{o}C for 4 hrs. Following restriction, the reactions were assessed by electrophoresis on 1% agarose gel and purified by using Geneaid gel/PCR DNA Fragments Extraction Kit (DF100). The purified fragments were ligated by preparing the following reaction mix:

Ligation mixture

Reagent	Volume (µl)
Excised *GmIPK2*_S fragment (100 ng)	2
Linearized pMINT (200 ng)	4
10X Ligase buffer	2
T4 DNA ligase (5 U/µl)	1
NFW	to 20

The ligation mixture was incubated at 4^{o}C overnight and used directly for the transformation of competent *E. coli* DH5α cells. The positive clones were identified by blue-white screening and confirmed by restriction digestion with *Bam*HI and *Cla*I to release ~700 bp GmIPK2_S plus GmFad2-1 fragment. The clone thus obtained was

named as pMIS.

4.2.3.4.2 Cloning of GmIPK2_As fragment into pMIS

Both pMAS and pMIS vectors were digested independently with *Cla*I and *Xho*I restriction enzymes in double digestion to facilitate directional cloning. The vector pMIS was first digested with the enzyme pair to linearize the plasmid with *Cla*I and *Xho*I overhangs. The pMS vector was then digested with the same enzyme pair to generate the same overhangs on the cloned GmIPK2_As PCR fragment. The GmIPK2_As with compatible cohesive ends was subsequently ligated into the linearized pMIS vector. The restriction digestions, DNA ligation and *E. coli* transformation were carried out as discussed in section 4.2.3.4.1. The positive clones were identified by blue-white screening and confirmed by restriction digestion with *Bam*HI and *Xho*I to release ~1 Kb *GmIPK2* ihp silencing cassette. The clone thus obtained was named as pMIHP.

4.2.4 Sub-cloning *GmIPK2* ihp cassette into plant expression vector pCWAK

The *GmIPK2* ihp construct thus generated was cloned to binary vector pCWAK, carrying seed specific vicilin promoter and *bar* gene selection marker, for further use of conducting soybean plant transformation. For sub-cloning silencing cassette into pCWAK, *GmIPK2* ihp cassette was excised from pMIHP by double digestion with *Bam*HI and *Xho*I and ligated to *Bam*HI and *Xho*I linearized pCWAK plasmid. The restriction of pMIHP was carried out as described in section 4.2.3.4.1. pCWAK was restricted in a reaction mixture containing:

Restriction mixture

Reagent	Volume (µl)
pCWAK (1-2 µg)	20
10X Buffer	5
*Bam*HI (20U/µl)	1
*Xho*I (10 U/µl)	2
NFW	to 50

The reaction mixtures were incubated at 37^{o}C for 4 hrs. Following restriction, the reactions were assessed by electrophoresis on 1% agarose gel and purified by using Geneaid gel/PCR DNA Fragments Extraction Kit (DF100). The purified fragments were ligated by adding insert and vector in 3:1 ratio respectively in the following reaction mix:

Ligation mixture

Reagent	Volume (µl)
Excised *GmIPK2* ihp fragment (150 ng)	3
Linearized pCWAK (50 ng)	1
10X Ligase buffer	2
T4 DNA ligase (5 U/µl)	1
NFW	to 20

The ligation mixture was incubated at 4^{o}C overnight and used directly for transforming competent *E. coli* DH5α cells. The positive clones were identified by blue-white screening and confirmed by restriction digestion with *Bam*HI and *Xho*I. The final clone thus obtained was named as pCWAK-IPK2ihp.

4.2.5 Mobilizing recombinant clones into *Agrobacterium EHA105*

The binary vector from confirmed recombinant pCWAK-IPK2ihp clones were mobilized into *Agrobacterium tumefaciens* strain EHA 105 by using a form of bacterial conjugation called triparental mating where the transfer is assisted by a

conjugative plasmid present in another bacterial strain. The vector pCWAK-IPK2ihp is capable of replicating in both *E. coli* as well as Agrobacterium. The transfer of pCWAK-IPK2ihp from the donor *E. coli* DH5α cells to the recipient *Agrobacterium* EHA 105 [397] cells was performed with the helper bacterial strain PRK 2013 in a seven days long procedure described below:

Day 1: 10 ml Luria-Bertani (LB) broth with antibiotic rifampicin (25 μg/ml) was inoculated with *Agrobacterium* EHA105 culture and incubated in a shaker incubator at 28^{o}C, 200 rpm for 46-48 hrs to obtain cells in their log phase (OD_{550} of 0.5-0.8).

Day 2: Both donor *E. coli* DH5α and helper PRK 2013 strains were inoculated in 10 ml LB broth with antibiotic kanamycin (50 μg/ml) for selection and incubated in a shaker incubator at 37^{o}C, 200 rpm for 16-18 hrs until the cells reach their log phase (OD_{600} of 0.5-0.6).

Day 3: Cultures of EHA 105, DH5α, and PRK 2013 were mixed in a ratio of 2:1:1 (400 μl: 200 μl: 200 μl respectively) in a sterile eppendorf tube under a laminar airflow hood, pipetted over three square Whatman filter paper blots (3mm in size) placed on a LA plate, and subsequently incubated at 28^{o}C for 46-48 hrs.

Day 5: Each blot was placed in 5 ml LB broth devoid of any antibiotic and mixed thoroughly. 100 μl of culture from each tube was diluted up to 1000 times using LB broth, spread on LA plates containing 25 μg/ml rifampicin & 50 μg/ml kanamycin, and incubated at 28^{o}C for 46-48 hrs.

Day 7: Single colonies were picked and cultured onto a master grid plate to identify positive clones harboring *GmIPK2* ihp cassette by colony PCR amplification of *Bar* gene:

Primers used for colony PCR amplification:

Bar FP: -5′ GAACGACGCCCGGCCGACAT 3′

Bar RP: -5′ GTCCAGCTGCCAGAAACCCAC 3′

Reaction mixture

Reagent	Volume (µl)
10X PCR buffer	2.5
dNTP mix (10mM each)	1
25mM $MgCl_2$	2
Bar FP (100 pmol/µl)	0.5
Bar RP (100 pmol/µl)	0.5
Template DNA (50ng/µl)	2
TaqDNA polymerase (5U/µl)	0.5
NFW	to 25

Amplification was performed by placing these PCR reaction mixtures on a PTC-100^{TM} Peltier thermal cycler (MJ Research).

PCR cycling protocol

Steps	Temperature	Time	Cycles
Initial denaturation	94^{o}C	4 min	1
Denaturation	94^{o}C	1 min	
Annealing	65^{o}C	30 sec	35
Extension	72^{o}C	30 sec	
Final extension	72^{o}C	4 min	1

On completion, 2µl of amplified product was analyzed by electrophoresis on 1% agarose gel along with a 1 kb DNA ladder to observe the amplicon size.

4.2.6 Transformation of soybean cotyledonary nodes and regeneration of transgenic plants

The protocol used for transformation and regeneration was standardized in our laboratory as follows:

4.2.6.1 Seed germination: Mature soybean seeds (*Glycine max* [L.] Merr. cv. Pusa-16) were surface sterilized with chlorine gas (mix 3.5 ml of 12 N HCl and 100 ml 4% sodium hypochlorite) in a desiccator under the fume hood for 16-18 hrs [398]. The sterilized seeds were placed on half gamborg's B5 basal medium [399], referred to as germination medium (GM) supplemented with 1 mg/l BAP, 3% sucrose, 0.6% agar, pH 5.8 and incubated for around for 5-6 days in a tissue culture chamber at 28°C, 16hrs/8hrs light/dark regime with an intensity of approximately 150 $\mu molm^{-2}s^{-1}$.

4.2.6.2 Explant preparation and co-cultivation: After 5-6 days, the germinated seedlings were taken out in an aseptic environment, their seed coat and radicle were removed to leave around 3-4 mm of its hypocotyl segment, and the seeds bisected by making a longitudinal cut through the cotyledonary node to separate it into two. The explants so derived were wounded around the nodal area with a scalpel blade and subsequently infected with *Agrobacterium* suspension culture for 30 min, prepared by suspending overnight grown *Agrobacterium* culture harboring pCWAK-IPK2ihp construct (OD_{600} 1-1.5) in liquid co-cultivation medium (CCM) containing 1/10 B5 salts supplemented with 1.68 mg/l BAP, 1 mM dithiothreitol (DTT), 200 μM acetosyringone, 1 mM sodium thiosulfate and 3.3 mM L-cysteine, pH 5.4. After infection, explants (8-10 per plate) were cultured with their adaxial side touching the Whatman filter paper lined solid CCM and incubated at 28^{o}C for 2-3 days in the dark [400-402]. Growth regulators, vitamins and antibiotics were prepared in autoclaved water and filter sterilized before use.

4.2.6.3 Shoot induction: The infected explants were then transferred adaxial side up with the hypocotyl embedded to full B5 solid shoot induction medium (SIM) supplemented with 1.2 mg/l BAP, 0.2 mg/l IBA, 200 mg/l cefotaxime, 100 mg/l

carbenicillin, 70 mg/l timentin, 50 mg/l kanamycin, 3% sucrose, 0.6% agar, pH 5.8 and incubated for 10-14 days at 28^{o}C, 16h/8h light/dark regime with an intensity of approximately 100 μmolm^{-2}s^{-1} for selective shoot induction. After two weeks, explants were sub-cultured to fresh SIM medium supplemented with 4 mg/l glufosinate (Glufosinate-ammonium, Sigma-Aldrich) for another 14 days.

4.2.6.4 Shoot elongation: Following shoot induction, the cotyledons were removed from the differentiating glufosinate resistant explants with newly developed shoots and sub-cultured to Murashige and Skoog (MS) medium [403] with B5 vitamins, referred to as shoot elongation medium (SEM). The SEM is supplemented with 0.75 mg/l GA_3, 0.1 mg/l IAA, 1 mg/l zeatin, 200 mg/l cefotaxime, 100 mg/l carbenicillin, 70 mg/l timentin, 100 mg/l kanamycin, 5 mg/l glufosinate, 3% sucrose, 0.6% agar, pH 5.8, and incubated for 10-14 days at 28^{o}C, 16hrs/8hrs light/dark regime with an intensity of approximately 100 μmolm^{-2}s^{-1}. The explants were sub-cultured to fresh SEM medium every two weeks until the shoot length becomes 2-3 cm in height.

4.2.6.5 Rooting: Explants with elongated shootlets were subsequently placed on half B5 rooting medium (RM) supplemented with 2 mg/l IBA, 200 mg/l cefotaxime, 100 mg/l carbenicillin, 70 mg/l timentin, 1% sucrose, 0.7% agar, pH 5.8 and incubated until roots are developed.

4.2.6.6 Hardening: Well rooted T_0 plantlets were transplanted to pot mixture containing vermiculite, cocopeat, and river sand in a 1:1:1 ratio and housed in tissue culture room for two to three weeks. Once acclimatized, the plants were shifted to National Phytotron Facility, IARI, Delhi, India, and grown to maturity at 28^{o}C under 16hrs/8hrs light/dark regime using 1000W high-pressure sodium lamps as the light source.

4.2.7 Screening and Evaluation of putative transgenic plants

4.2.7.1 Confirmation of T_0 positive transformants by PCR

Glufosinate-resistant plants were screened for the presence of *IPK2* ihp construct in their genomic DNA by PCR amplification using *bar* gene and ihp construct-specific primers. Genomic DNA was isolated from the leaves of putative positive transformants and non-transformant plants using the protocol described by Murray and Thompson [404]. 5 g of leaves tissue was finely powdered in liquid nitrogen, transferred to a 50 ml centrifuge tube containing pre-warmed DNA extraction buffer (0.7 M NaCl, 1% CTAB, 50 mM Tris-HCl (pH 8.0), 10 mM EDTA, and 1% 2-mercaptoethanol), and incubated at 65^{o}C in a water bath for an hour. An equal volume of chloroform:isoamyl alcohol (24:1) was added to the tube, mixed well, and centrifuged at 10,000 g for 10 min at RT. The upper aqueous layer was carefully transferred to a fresh sterile tube, mixed with one volume of chilled isopropanol, and incubated at RT for at least an hour. The precipitated DNA was spooled out or pelleted down by centrifugation at 10,000 g for 10 minutes at 4^{o}C. The pellet was washed with 70% ethanol, dried, and dissolved in TE buffer, pH 8.0 (10 mM Tris-HCI, pH 8.0 and 1mM EDTA, pH 8.0). Isolated DNA was purified by treating it with 20 µg/ml DNase free RNase A, followed by phenol:chloroform:isoamyl alcohol (25:24:1) extraction. The DNA was precipitated with a 1/10th volume of 3M sodium acetate (pH 5.2) & 2.5 volume of ice-cold absolute ethanol and stored at -20^{o}C. The purified DNA fragment was checked on 1% agarose gel for integrity. About 100 ng of the above-isolated DNA was used to check for the presence of the *bar* gene by carrying out amplification as described under section 4.2.5. The presence of a fragment of ihp cassette was also verified by following amplification protocol described below:

Primers used for amplification:

GmIPK2_S FP: -5′ CGCGGATCCGCGTTGCAGAAGCTCAAG 3′

GmFad2-1 RP: -5' CCATCGATGGTATACCGCACTAGTAAACCAC 3'

Reaction mixture

Reagent	Volume (µl)
10X PCR buffer	2.5
dNTP mix (10mM each)	1
25mM $MgCl_2$	2
GmIPK2_S FP (100 pm/µl)	0.5
GmFad2-1 RP (100 pm/µl)	0.5
Genomic DNA (100 ng)	3
TaqDNA polymerase (5U/µl)	0.5
NFW	to 25

Amplification was performed by placing these PCR reaction mixtures on a PTC-100^{TM} Peltier thermal cycler (MJ Research).

PCR cycling protocol

Steps	Temperature	Time	Cycles
Initial denaturation	94^{o}C	4 min	1
Denaturation	94^{o}C	30 sec	35
Annealing	62^{o}C	30 sec	
Extension	72^{o}C	30 sec	
Final extension	72^{o}C	10 min	1

On completion, 2µl of amplified product was analyzed by electrophoresis on 1% agarose gel along with a 1 kb DNA ladder to observe the amplicon size.

4.2.7.2 Analysis of PCR confirmed transformants by southern blotting

Southern hybridization analysis was carried out on T_1 plants cultivated from the seeds of PCR positive T_0 lines to confirm stable T-DNA integration and to estimate the number of copies inserted.

4.2.7.2.1 Restriction digestion of genomic DNA: 20 µg of genomic DNA from each sample was digested with *Pst*I which does not cut the T-DNA region by preparing the reaction mixture as follows:

Restriction mixture

Reagent	Volume (µl)
genomic DNA (2 µg)	10
10X NEBuffer	2.5
*Pst*I (20 U/ µl)	5
NFW	to 25

The reaction mixture was incubated at 37^{o}C overnight. *Xho*I digested pCWAK-IPK2 plasmid DNA was used as positive control. The restricted samples were separated by electrophoresis on 0.8% agarose gel.

4.2.7.2.2 Transfer of DNA on to nylon membrane: The gel was removed from the electrophoresis unit and transferred onto a positively charged nylon membrane (Axiva, India) by capillary blotting following the method described by Sambrook and co-workers [278]. After visualization, the gel was transferred in a glass tray and soaked in 0.25 M HCl for 15 min with slow shaking at RT to carry out depurination. The DNA was denatured by soaking the gel in several volumes of denaturing solution (1.5 M NaCl and 0.5 M NaOH) for 1 hr followed by its neutralization in 0.5 M Tris-HCl pH 7.2 and 1.5 M NaCl solution for 1 hr with slow shaking at RT. A piece of 3MM Whatman filter paper was wrapped around a glass plate and placed inside a glass reservoir filled with 20X SSC (3M NaCl and 0.3 M Sodium citrate) such that both its ends dip in SSC to serve as a wick. The filter paper was smoothened with a glass rod to remove all the air bubbles trapped between the paper and the glass plate. The gel was then placed on the filter paper in an inverted manner so that the wells are now facing the glass plate and again smoothened to remove all the air bubbles. A

nylon membrane (1-2 MM larger than gel size) saturated with 20X SSC was placed on the gel to make perfect contact avoiding any air bubbles. Two pieces of exactly same size Whatman filter paper were placed on the membrane and again smoothened to remove all the air bubbles. A 6 cm stack of absorbent paper towels was placed on top of it and weighed down by placing a weight of 500 g over it. The whole assembly was covered with waterproof saran wrap to prevent short-circuiting of fluids between the paper towel and the Whatman paper placed under the gel. The transfer of DNA was allowed to proceed for 12-16 hr. The paper towels and Whatman filter were then carefully removed, and the position of the wells was marked on the nylon membrane using a soft pencil. The membrane was peeled and washed with 5X SSC to remove any traces of the gel and dried properly at RT. The DNA strands thus transferred were cross-linked to the nylon membrane by exposure to UV in a crosslinker for 2 min and stored in a sheet of filter paper at RT for further use.

4.2.7.2.3 Pre-hybridization and labelling of the probe: The cross-linked membrane was soaked in pre-hybridization buffer and placed carefully into a clean hybridization bottle containing 20 ml (0.2 ml cm^{-2} of the blot) of pre-warmed (65^oC) pre-hybridization buffer (5x Denhardt's solution, 5X SSC, 50mM phosphate buffer, 0.1% SDS, 50% formamide, 0.5 µg sheared salmon sperm DNA). The blot was incubated for 2h at 65^oC in hybridization chamber. For probe preparation, 500 bp *bar* gene fragment was PCR amplified and gel purified as described under the section 4.2.5. The purified fragment was biotin labelled using Biotin DecaLabel DNA Labeling Kit (Thermo Scientific, USA) by setting up a reaction described by the manufacturer as follows:

Reagent	Volume (µl)
PCR amplified probe DNA (150 ng)	5
Decanucleotide in 5X reaction buffer	10
NFW	to 44

The contents of the tube were mixed, incubated in a boiling water bath for 5-10 min and cooled on ice. To the same tube, the following components were added:

Reagent	Volume (µl)
Biotin labelling mix	5
Klenow fragment, exo$^-$ (5 U)	1

The above components were mixed and incubated for 1 hr at 37^{o}C. The reaction was stopped by the addition of 1 µl of 0.5 M EDTA, pH 8. The labeled DNA was used directly for hybridization.

4.2.7.2.4 Hybridization and detection: After 2 hrs of pre-hybridization, the biotin-labelled probe was added to the pre-hybridization bottle. Hybridization of the probe was carried out overnight by incubation at 65^{o}C in the hybridization chamber. After hybridization, the membrane was washed and detected using Biotin Chromogenic Detection Kit according to the protocol described by the manufacturers (Thermo Scientific, USA). Hybridization solution was discarded, and the blot was washed thrice with 2X SSC and 0.1% SDS for 15 min each at 50^{o}C with moderate shaking. To further increase the stringency, the blot was washed twice with 1X SSC and 0.1% SDS and again twice with 0.1X SSC and 0.1% SDS each at 50^{o}C for 5 min. In the final washing, the blot was rinsed in 30 ml of blocking buffer provided in the detection kit for 5 min at RT with moderate shaking. The membrane was blocked by shaking it in 30 ml of blocking solution for 30 min at RT. 20 ml of the 5000-fold diluted streptavidin-AP conjugate was subsequently incubated with the membrane for

30 min at RT with moderate shaking to allow streptavidin-biotin bond formation. After 30 min, the membrane was washed twice with 60 ml of blocking buffer for 15 min and incubated with 20 ml of detection buffer for 10 min. The membrane was detected by incubation in 10 ml of freshly prepared substrate solution at RT in the dark to perform the enzymatic reaction. The blue-purple precipitate becomes visible after 15-30 min of incubation. The membrane was rinsed with sterilized double distilled water for a few seconds to stop the reaction. The blot was finally removed from the water, air dried and documented on syngene gel documentation system.

4.2.7.3 Quantitative expression analysis of transgenic progenies

Southern positive T_2 plants were further analyzed by semi-quantitative RT-PCR and quantitative real-time PCR (qRT-PCR) to estimate *GmIPK2* transcript levels. Total RNA was isolated from fresh soybean seed (8-10 mm) tissue of T_2 transgenic and non-transgenic plants using TRI Reagent (Invitrogen, USA) and first-strand cDNA was synthesized from 5 μg of DNase I-treated total RNA using RevertAid^{TM} H Minus First strand cDNA Synthesis Kit (Thermo Scientific, USA) by following the protocol described under section 3.2.2.1.

4.2.7.3.1 RT-PCR expression analysis

The *GmIPK2_S* fragment was subsequently amplified from the cDNA template following the exact procedure described under section 4.2.3.1 along with the amplification of house-keeping gene phosphoenolpyruvate carboxylase (*PEPCo*) (qPEPCoF5'-CATGCACCAAAGGGTGTTTT-3', qPEPCoR 5'-TTTTGCGGCAGCT ATCTCTC-3') which was used as the endogenous control. PCR amplification of *PEPCo* was carried by following the same cycling protocol except for its annealing

which was carried out at 60^{o}C. On completion, the amplification products were analyzed by electrophoresis on 1% agarose gel alongside a 1 kb molecular weight marker.

4.2.7.3.2 Quantitative real-time PCR analysis

Quantitative real-time PCR analysis (qRT-PCR) analysis was subsequently performed based on SYBR green chemistry using DyNAmo Flash SYBER Green qPCR Kit, Thermo Scientific, USA. The amplification was performed on PikoReal 96 Real-Time PCR platform (Thermo Scientific, USA) using gene-specific primers, designed to amplify a 124 bp fragment from its conserved region (qIPK1F 5'-GGAGCGCT TGCAGAAGC-3', qIPK1R5'-GACCAGAGGGTTGGTAGC-3'). To normalize the variance among samples, house-keeping gene *PEPCo* was used as the endogenous control. The experimental reactions were prepared by adding components in the following order:

qPCR reaction mixture

Reagent	Volume (µl)
2X Master mix	10
Primer Mix (0.25 µM each)	2
cDNA template (5 ng/µl)	4
NFW	to 20

The 2X master mix used in the reaction mixture contains a hot-start version of a modified *Tbr* DNA polymerase, SYBR® Green I, optimized PCR buffer, 5mM $MgCl_2$, dNTP mix including dUTP. Amplification was then achieved by performing a two-step PCR reaction with an initial denaturation at 94°C for 4 min followed by 40 cycles of denaturation at 94°C for 30s, annealing/extension at 60°C for 30s, and fluorescence data collection. Melt curve analysis was performed from 60-95°C to

determine the specificity of each unique amplification product. To minimize variation in the output, we considered three technical replicates corresponding to each of the three biological replicates. The relative abundance of *GmIPK2* was calculated using the 2-ΔΔCT method [281].

4.2.7.4 Analysis of phytic acid concentration by HPLC

We further estimated PA concentration in transgenic seeds by performing reversed-phase high-performance liquid chromatography (RP-HPLC) as described by Lehrfeld [405] with some modifications [406].

4.2.7.4.1 Sample preparation: T_3 transgenic seeds were dried at 37°C till a constant weight was achieved and ground in a Tecator Cyclotec 1093 Sample mill. 500 mg of flour was then defatted overnight in 10 ml hexane and subsequently extracted with 2.4% HCl by sonication for 3 mins (QSonica, USA) followed by mechanical agitation for 1 hr at 250 rpm, 37°C. The extract was centrifuged (Thermo Scientific, USA) at 13,000 g for 20 min and one part of clear supernatant recovered was mixed with four parts of HPLC grade water. The diluted sample passed through a conditioned SAX column (Hypersep, Thermo Scientific, USA) connected to a vacuum manifold (Millipore, USA at 75-100 mmHg), was washed with 2 ml each of HPLC grade water and 5 mM HCl, and finally eluted with 5 ml of 2 M HCl to separate the inositol phosphates. The filtrate was evaporated in a vacuum rotary evaporator (Hei-VAP Value Digital G3, Heidolph Germany) at 40°C/80 rpm, re-dissolved in 1 ml of mobile phase (acetonitrile, formic acid and tetrabutylammonium hydroxide, Merck, USA; 4.8:5.1:0.1, v/v/v), and filtered using a 0.22 μm PVDF syringe filter (Millipore, USA). PA dilutions 0.0625 to 4.0 mg/ml were prepared in the mobile phase, filtered, and injected into the HPLC column to construct a calibration curve. All the reagents used

for extraction and quantification of PA were of HPLC grade.

4.2.7.4.2 Phytic acid estimation: For PA HPLC analysis, we used a Shimadzu HPLC system equipped with a binary pump, online degasser, a Rheodyne injection valve, and a column oven. We injected the sample onto a silica-based reversed-phase C_{18} column (250×4.6 mm, 5µ; Shimadzu, Japan) equilibrated with isocratic mobile phase at a flow rate of 1.0 ml min^{-1}. The temperature of the column oven was set at 40°C to avoid hydrolysis of PA to inositols and ortho-phosphates at higher temperatures [407] as well as to reduce back pressure. PA signals were monitored with a UV-VIS photodiode array detector (SPD-M20, Shimadzu, Japan) coupled to the solvent delivery system by taking readings at a wavelength of 197 nm. The detector signals were processed and integrated by the data acquisition system, and the final PA concentration was calculated using the calibration curve prepared with a PA dodeca-sodium salt standard (Sigma Aldrich, USA). The software used for analysis was LC-Solutions (version 1.25).

4.2.7.5 Estimation of seed phosphorus levels

We then estimated the total phosphorus (total P) and inorganic phosphorus (P_i) levels in transgenic as well as non-transgenic seeds to correlate it with the observed reduction in PA content. Total P in the seeds was determined colorimetrically by a method described by Chen and co-workers [408]. We prepared the samples (transgenic and non-transgenic control) for the assay by wet ashing 250 mg dried (for 48 hrs at 60^{o}C) ground seeds (in Tecator Cyclotec 1093 Sample Mill) in 2 ml of concentrated sulfuric acid [150, 157]. 1 ml of each sample was mixed with 1 ml of Chen's reagent (10% ascorbic acid, 3 M sulphuric acid, 2.5% ammonium molybdate and water in 1:1:1:2 ratio) and incubated at 37^{o}C for 1.5 hr. Their absorbance was

measured at 882 nm, and the total P in the samples was estimated by preparing a standard curve of sodium phosphate dibasic heptahydrate at the concentration of 0, 0.17, 0.34, 0.85, 1.70, 3.40 µg.

For the determination of Pi levels, we followed a protocol described by Amery and co-workers [409] which is a modified Chen's method that uses single seed chips (1-2 mg) derived from cotyledonary segment opposite to the embryonic axis for estimation. Seed samples were extracted using 50 µl buffer (25 mM magnesium chloride and 12.5 % trichloroacetic acid) for 14-16 hrs at 37^{o}C with gentle shaking. 10 µl sub-sample of extracts were loaded in triplicates onto a 96 well plate and diluted with 90 µl water. Each sample was incubated with 100 µl of Chen's reagent at 37^{o}C for 1 hr. Final absorbance of the samples was read at 882 nm on a GloMax®-Multi Detection System (Promega, USA) and their P_i was estimated from a standard curve prepared by using sodium phosphate dibasic heptahydrate at the concentrations of 0, 0.17, 0.34, 0.85, 1.70, 3.40 µg.

4.2.7.6 Bioavailability assay using *in vivo* simulation model

To determine iron (Fe), zinc (Zn) and calcium (Ca) bioavailability, we milled (Tecator Cyclotec 1093 Sample Mill) the mature seeds (transgenic and non-transgenic) and suspended 1g of each flour sample in 10 ml distilled water in triplicates for digestion under simulated gastrointestinal conditions following the method described by Kiers and co-workers [410]. We subjected the samples to enzymatic degradation by sequential incubation with α-amylase solution (see Appendix II) for 30 min at 37°C, stomach medium (see Appendix II) for 1 hr at 37°C, and 2% pancreatic solution (see Appendix II) for 30 min at 37°C. The digest was centrifuged at 9000 rpm for 15 mins at 4°C and the supernatant obtained was passed through a 0.45 mm filter. Distilled water was

processed similarly to use as blank for sample correction. Filtered supernatants from both, samples and blank were analyzed for Fe, Zn, and Ca using Atomic Absorption Spectrophotometer (AAnalyst 200, Perkin Elmer, USA) and the % bioavailability was calculated as follows:

$$\% \text{ Bioavailability} = \frac{\text{Amount of Fe, Zn, Ca (sample)} - \text{Amount of Fe, Zn, Ca (blank)}}{\text{Amount of Fe, Zn, Ca in the undigested sample}} \times 100$$

4.2.7.7 Agronomic evaluation of transgenic plants

4.2.7.7.1 Germination assay for seed viability

We assessed the germination capacity of T_3 seeds in comparison to non-transgenic control by performing a controlled germination test (CGT) described by Campion and co-workers [411]. The seeds were immersed under water for 8 hrs at 28°C and then transferred to fresh water for an additional period of 12 hrs. Following incubation, seeds were rinsed three to four times in distilled water and subsequently germinated in Petri dishes lined with filter paper soaked in distilled water under a photoperiod of 8 hrs dark and 16 hrs light at 28°C. We also monitored seed vigor by using the standard accelerated aging test (AAT) test. In AAT, we subjected unimbibed seeds to high temperature (42°C) and relative humidity (90%) for 72 hrs. We then removed the seeds from stress conditions, rinsed them three to four times in distilled water and placed them on Petri dishes lined with filter paper soaked in distilled water for germination under an optimum temperature of 28°C and a photoperiod of 8 hrs dark and 16 hrs light.

4.2.7.7.2 Phenotypic analysis

To capitulate the influence of transgene integration on phenotype, we evaluated

different morphological traits of transgenic plants grown under controlled environmental conditions. We randomly chose five matured plants from transgenic lines P2-45 and P6-39 and compared their growth parameters with non-transgenic control. We determined the plant height, number of pods, number of seeds, seed dry weight, stem length, and root length. To calculate mean seed dry weight (SDW), we harvested up to 100 matured seeds and dried them in the oven at 60°C for 72-96 hrs before weighing them.

4.3 Results and Discussion

4.3.1 Designing *GmIPK2* ihp construct

To design *GmIPK2* ihp construct, *GmIPK2_S* and *GmIPK2_As* fragments each ~305 bp in length and *GmFad2-1* intronic fragment ~396 bp in length were amplified using their respective primers with suitable restriction enzyme sites and cloned individually in pCR™2.1-TOPO® TA cloning vector as described under section 4.2.3 (Figure 4.2).

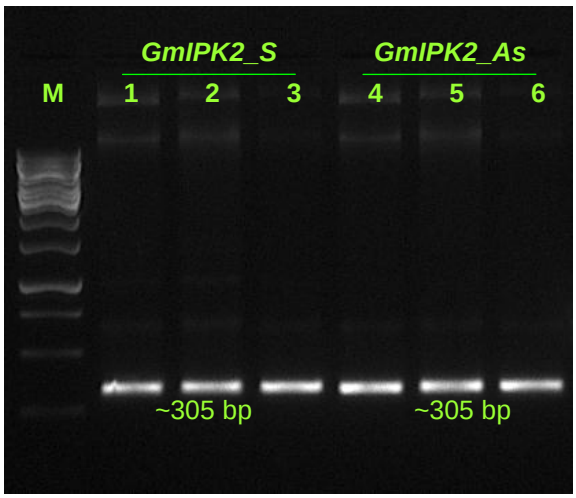


Figure 4.2: (A) PCR amplification of *GmIPK2 _S* and *GmIPK2 _As* fragments. Lane M: O'Gene Ruler™ 1 Kb DNA ladder, lanes 1-3: ~305 bp *GmIPK2_S* amplicon (T_m=65.3°C), lanes 4-6: ~305 bp *GmIPK2_As* amplicon (T_m=65.3°C)

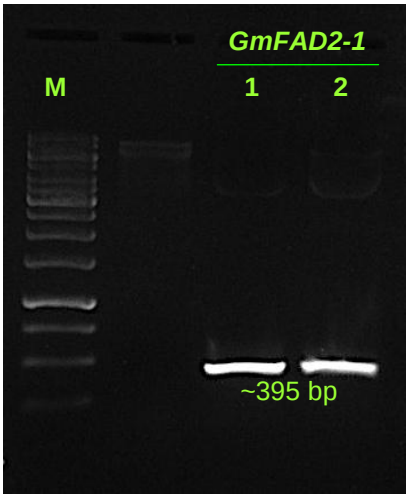
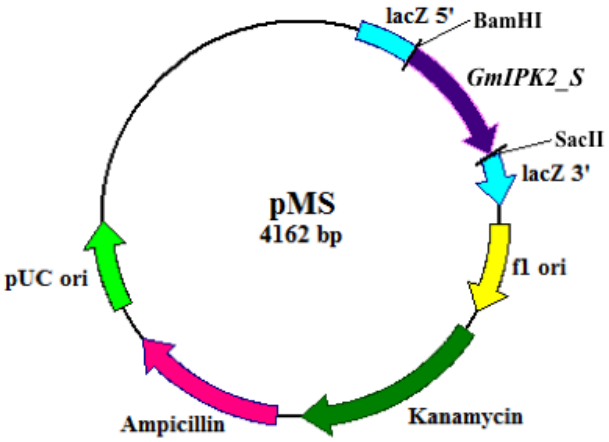


Figure 4.2: (B) PCR amplification of *GmFAD2-1* intron. Lane M: O'GeneRuler^{TM} 1 Kb DNA ladder, lanes 1,2: ~395 bp *GmFAD2-1* amplicon (T_m=65.5^oC).

We confirmed the positive pMS (Figure 4.3), pMAS (Figure 4.4), and pMINT clones by EcoRI restriction digestion. Electrophoresis on 0.8% agarose gel determined the size of the expected fragments. Since pMINT was used as the base vector for assembling the final silencing construct, we confirmed the orientation of the intronic insert by double digestion with *Cla*I/*Nco*I to obtain restriction fragments of expected sizes (Figure 4.5). Once orientation of the intron was confirmed, we prepared the *GmIPK2* ihp silencing cassette by ligating the sense (Figure 4.6) and antisense fragments with the intronic fragment in the base vector as per the procedure described under section 4.2.3. We identified the positive clones by performing double digestion with *Bam*HI and *Xho*I to release ~1 Kb *GmIPK2* ihp cassette (Figure 4.7) and confirmed them by DNA sequence analysis of overlapping plasmid clones using M13 universal primers on an automated sequencer (ABI 3730xl DNA Analyzer, USA) (Figure 4.8). The confirmed clones were named as pMIHP. The ihp cassette thus generated was sub-cloned to binary vector pCWAK, carrying seed specific vicilin promoter and *bar* gene selection marker, for further use in conducting soybean plant transformation. The positive clones were confirmed by restriction analysis with

different enzymes (Figure 4.9) and named as pCWAK-IPK2ihp (Figure 4.10).

(A)



(B)

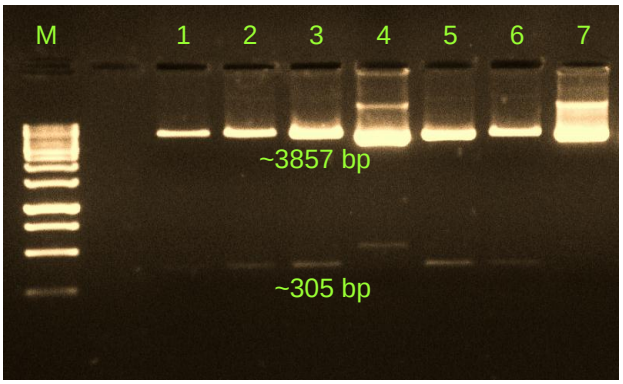
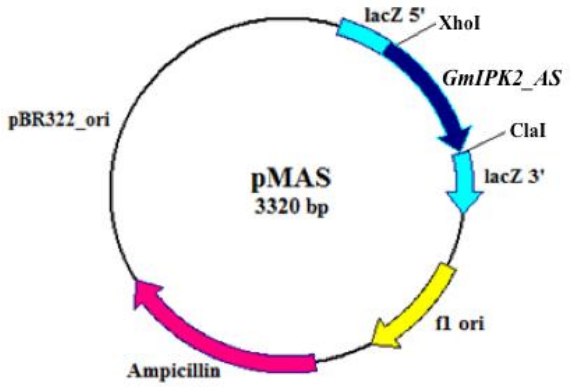


Figure 4.3: (A) Circular map of pMS vector containing *GmIPK2_S* cloned in pCR™2.1-TOPO cloning vector. **(B)** Restriction analysis of pMS clones with *Eco*RI. Lane M: O'GeneRuler™ 1 Kb DNA ladder, lanes 3 and 5: positive clones releasing ~305 bp *GmIPK2_S* fragment and ~3857 bp pCR™2.1-TOPO backbone.

(A)



(B)

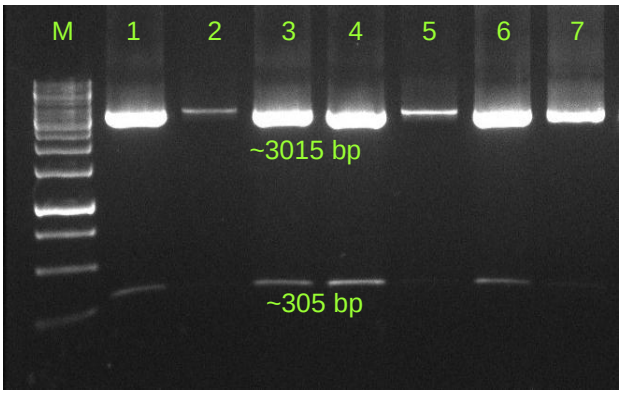
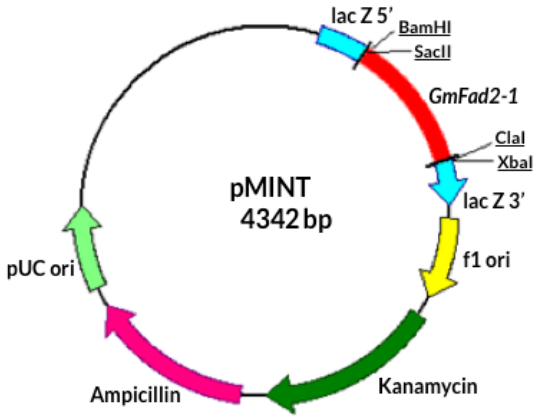


Figure 4.4: (A) Circular map of pMAS vector containing *GmIPK2_AS* cloned in pGEM®-T Easy cloning vector. **(B)** Restriction analysis of pMAS clone with *Eco*RI. Lane M: O'Gene Ruler^{TM} 1 Kb DNA ladder, lanes 1-7: *Eco*RI digested clones, lanes 1, 3, 4, 6, and 7: positive clones releasing ~305 bp *GmIPK2_AS* fragment and ~3015 bp pGEM®-T Easy vector backbone.

(A)



(B)

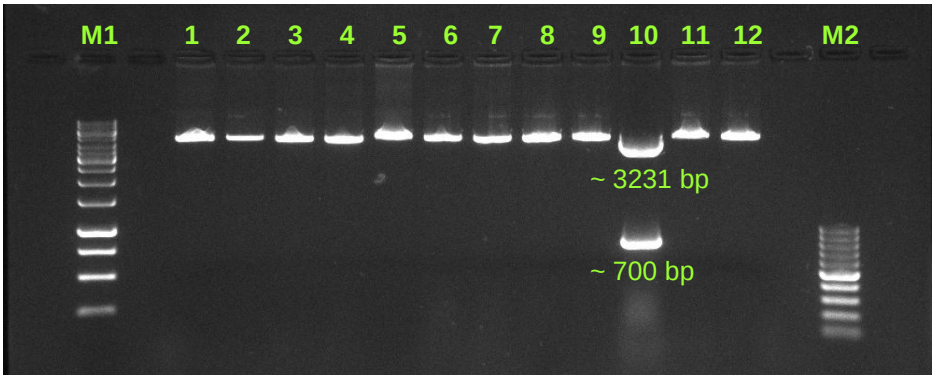


Figure 4.5: (A) Circular map of the pMINT vector containing *GmFad2-1* cloned in a pCR™2.1-TOPO cloning vector. **(B)** Restriction analysis of pMINT clones with *Cla*I and *Nco*I. Lane M1: O'GeneRuler™ 1 Kb DNA ladder, Lane M2: Thermo Scientific™ λDNA/*Eco*RI+*Hin*dIII marker, lanes 1-7: clones with undesirable intron orientation, lanes 8-10: clones showing desired orientation of the intron confirmed by size of the fragments obtained.

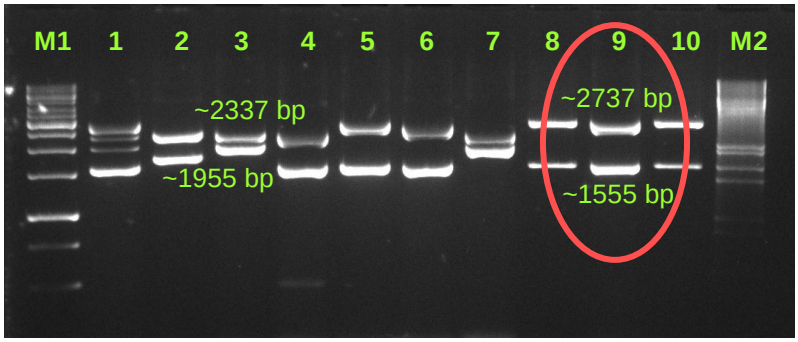


Figure 4.6: Restriction analysis of pMIS clones with *Bam*HI and *Cla*I. Lane M1: O'Gene Ruler^{TM} 1 Kb DNA ladder, lane M2: O'GeneRuler^{TM} 100 bp DNA ladder, lane 10: positive pMIS clone releasing ~700bp *GmFad2-1* + *GmIPK2_S* fragment.

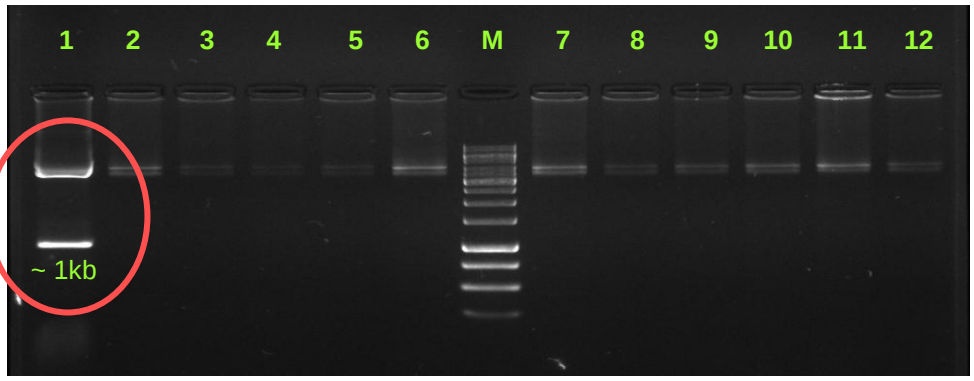


Figure 4.7: Restriction analysis of pMIHP clones by double digestion with *Bam*HI and *Xho*I. Lane M: O'GeneRuler^{TM} 1 Kb DNA ladder, lane 1: positive pMIHP clone releasing ~1 Kb *GmIPK2* ihp cassette.

Figure 4.8: The sequence of *GmIPK2* ihp cassette derived by analysis on an automated sequencer using M13 universal primers. The conserved 5' and 3' splice sites are shown flanking the *GmFad2-1* intron.

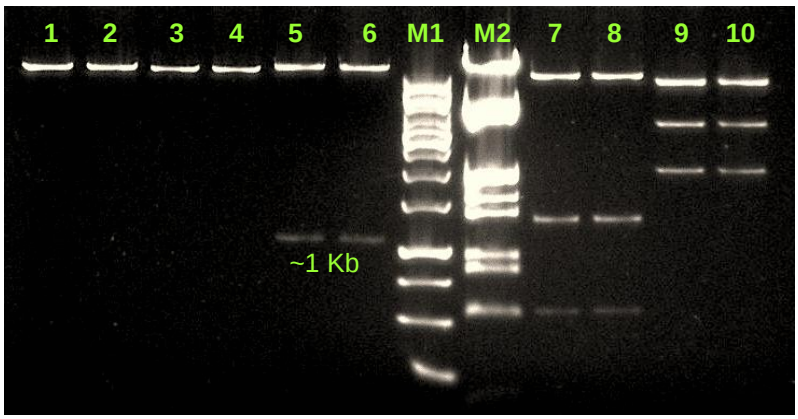


Figure 4.9: Restriction analysis of two pCWAK-IPK2ihp clones with different enzymes. Lane M1: O'GeneRuler™ 1 Kb DNA ladder, lane M2: Thermo Scientific™ λDNA/*Eco*RI marker, lanes 1, 2: digestion with *Bam*HI generating linearized plasmid, lanes 3, 4: digestion with *Xba*I also generating linearized plasmid, lanes 5, 6: double digestion with *Bam*HI and *Xho*I releasing ~1 Kb *GmIPK2* ihp cassette, lanes 7, 8: digestion with *Xho*I releasing three bands, lanes 9, 10: digestion with *Hin*dIII also releasing three bands. Restriction of both the pCWAK-IPK2ihp clones with each enzyme generates expected number and size of fragments.

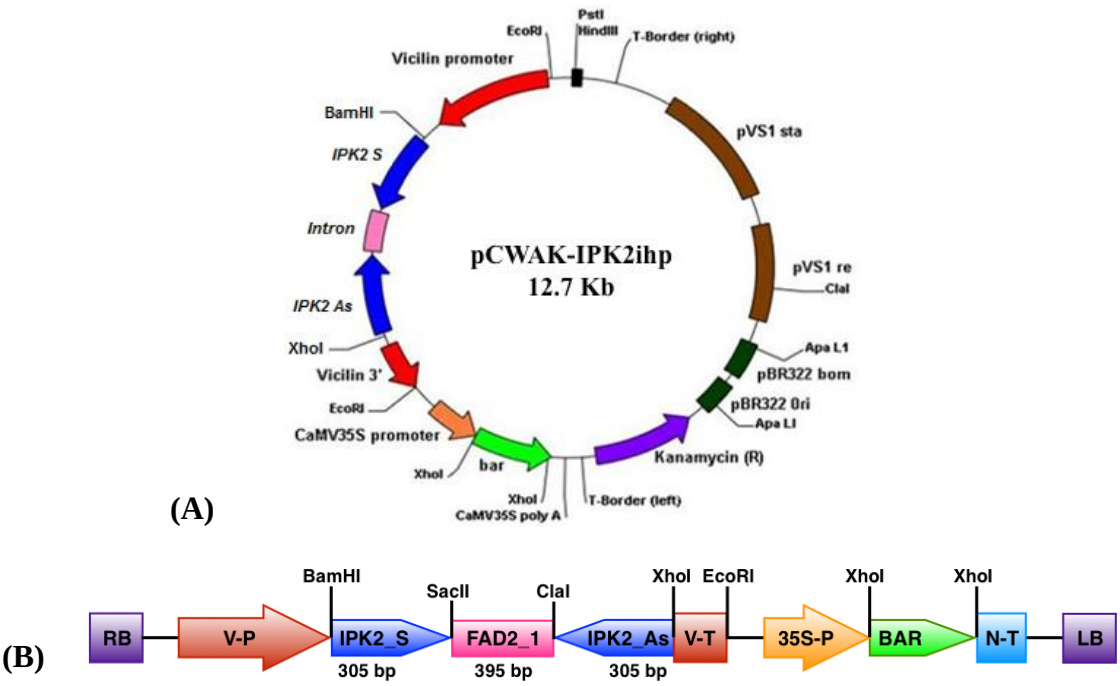


Figure 4.10: (A) Circular map of the binary vector construct pCWAK-IPK2ihp harboring the vicilin-GmIPK2ihp expression cassette for soybean transformation. **(B)** Linear map of the T-DNA region of pCWAK-IPK2ihp. RB: right border; V-P: vicilin promoter; IPK2_S: *IPK2* sense fragment; FAD2_1: fatty acid oley1 Δ12 desaturase gene intron; IPK2_AS: *IPK2* antisense fragment; V-T: vicilin terminator; 35S-P: cauliflower mosaic virus 35S promoter; BAR: bialaphos resistance gene; N-T: nopaline synthase terminator; LB: left border.

4.3.2 Transformation of pCWAK-IPK2ihp binary construct into soybean

The binary vector pCWAK-IPK2ihp was mobilized into *Agrobacterium tumefaciens*

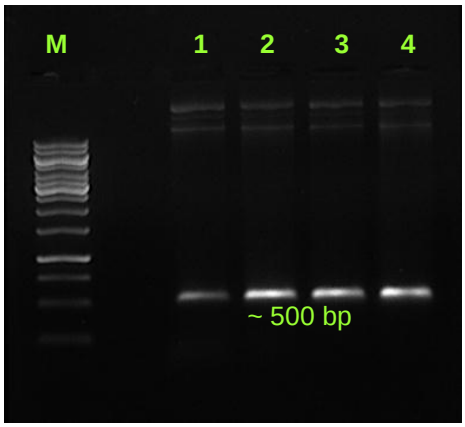


Figure 4.11: PCR amplification of *bar* gene to identify *A. tumefaciens* strain EHA105 clones harboring binary vector pCWAK-IPK2ihp construct. Lane M: O'GeneRuler^{TM}

1 Kb DNA ladder, lanes 1-4: positive clones showing ~500 bp *bar* gene amplicon.

strain EHA105 by triparental mating. Agrobacterium colonies harboring pCWAK-IPK2ihp vector were identified by colony PCR amplification of ~500 bp *bar* gene amplicon (Figure 4.11). The positive colonies were used for *Agrobacterium*-mediated cotyledonary node transformation of pCWAK-IPK2ihp construct into soybean. The cotyledonary node explants derived from 5-6 days old seedlings of soybean cv. Pusa-16 (Figure 4.12) were infected with *Agrobac-terium* culture harboring binary vector pCWAK-IPK2ihp (Figure 4.13) as discussed under section 4.2.6.2.

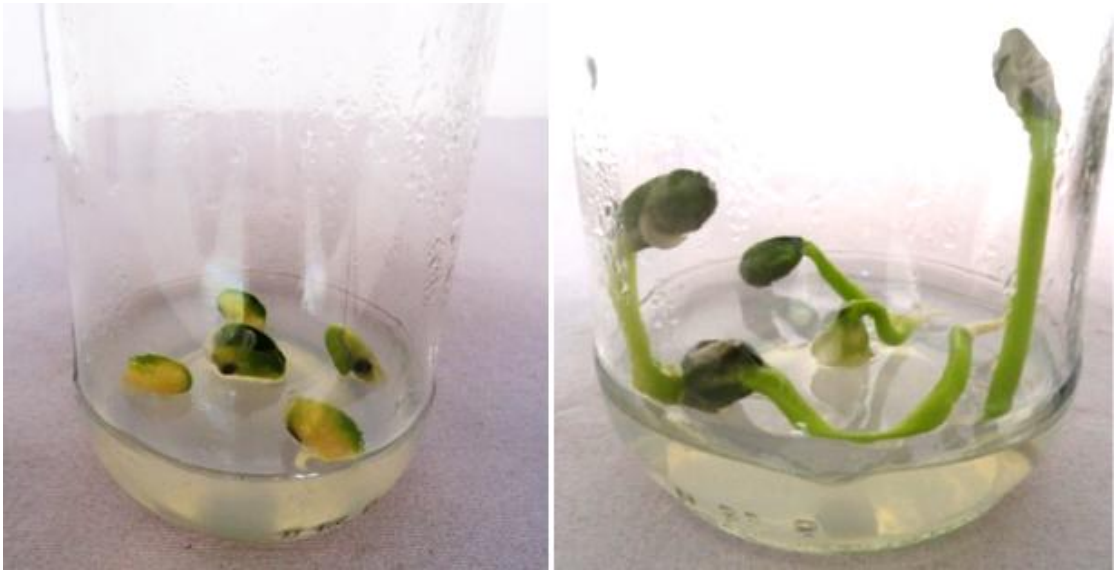
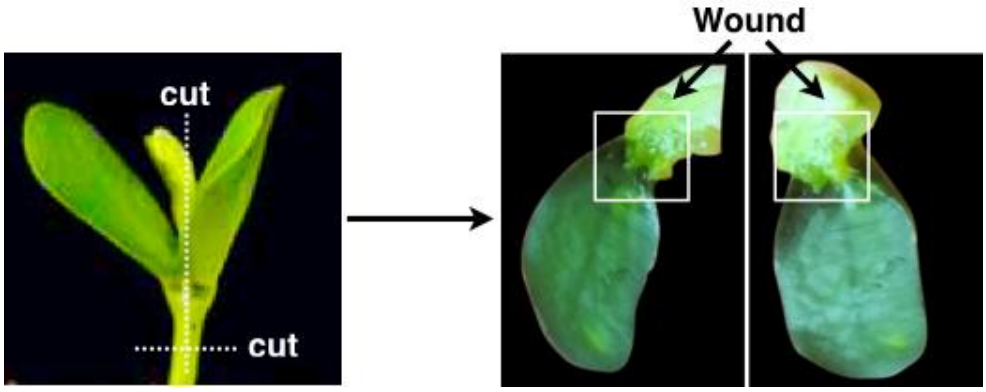


Figure 4.12: Germination of sterilized soybean Pusa-16 seeds in 1/2 B5 media up to 5-6 days for explant preparation.

(A)



(B)

Figure 4.13: (B) Wound infection by recombinant *Agrobacterium* harboring pCWAK-IPK2ihp vector and subsequent incubation of cotyledons on solid co-cultivation media for 3-4 days.

Explants were then sub-cultured on a non-selective SIM media to initiate shoot development (Figure 4.14) and further sub-cultured to fresh SIM medium supplemented with 4 mg/l glufosinate for another 14 days (Figure 4.15). All the non-transformed, glufosinate sensitive tissues showed senescence and subsequent necrosis. Four weeks later, healthy explants were transferred to SEM containing 5 mg/l glufosinate for continued selection (Figure 4.16) and sub-cultured in fresh SEM every two weeks until shootlets approximately 2-3 cm high were developed (Figure 4.17). Rooting was subsequently induced in elongated shootlets by the application of auxin

phytohormone IBA, supplemented in RM (Figure 4.18). After 20-25 days, well-rooted plantlets were transplanted to soil pots for hardening (Figure 4.19) and grown to maturity under 16hrs/8hrs light/dark regime at National Phytotron Facility, IARI, Delhi, India (Figure 4.20).

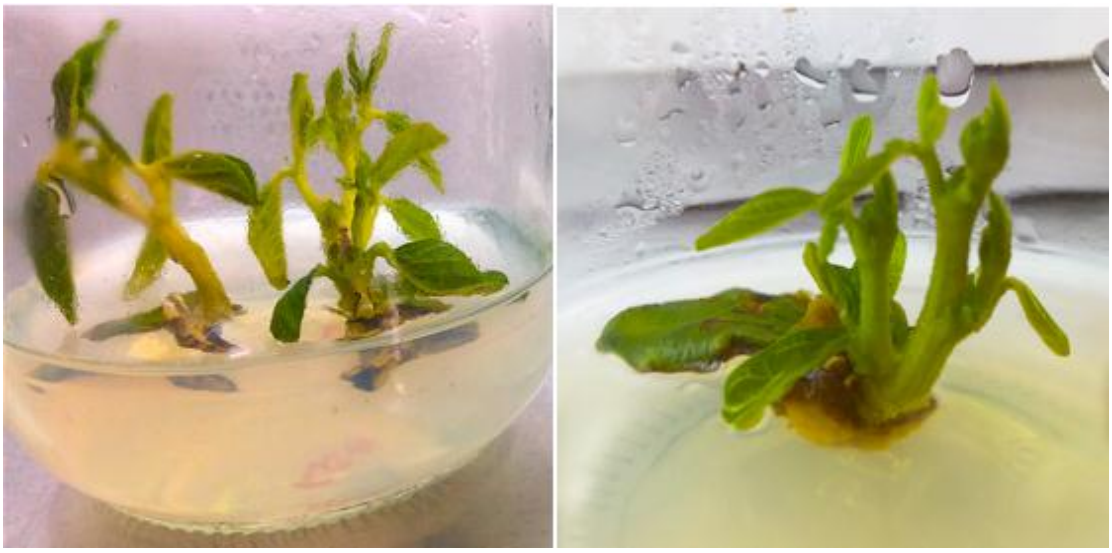


Figure 4.14: Multiple shoot induction from *agrobacterium* co-cultivated explants after 14 days of culture on shoot induction medium (SIM-I) without glufosinate selection.

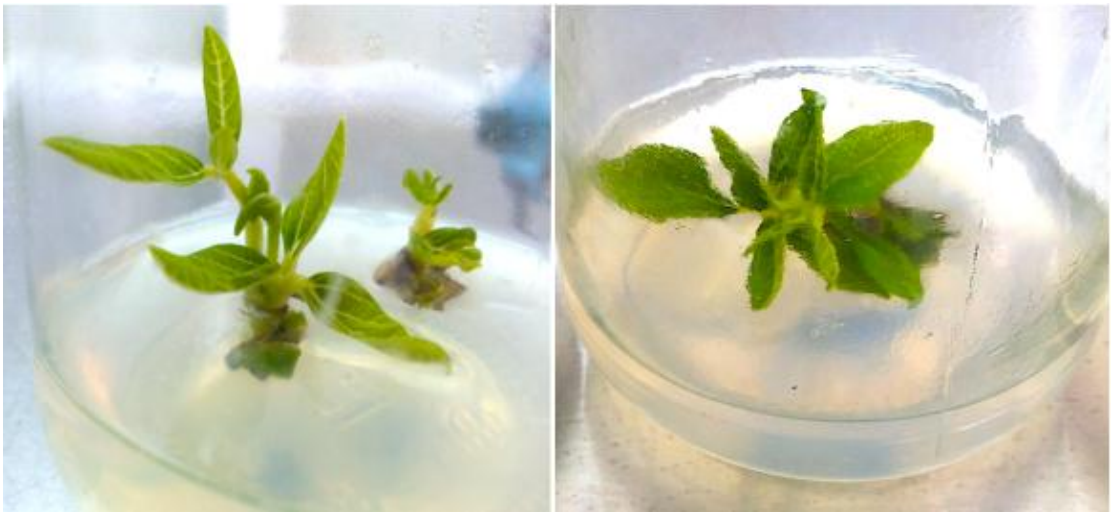


Figure 4.15: Selection of adventitious shoots on shoot induction medium (SIM-II) with 4 mg/l glufosinate after 28 days of culture initiation.

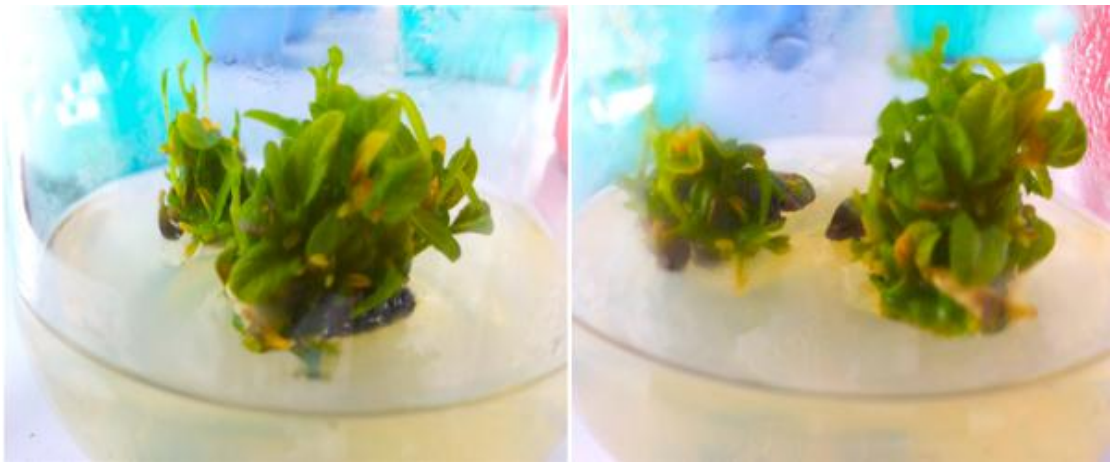


Figure 4.16: Multiple shoot proliferation from explants on shoot elongation medium (SEM-I) with 5 mg/l glufosinate after 42 days of culture initiation.

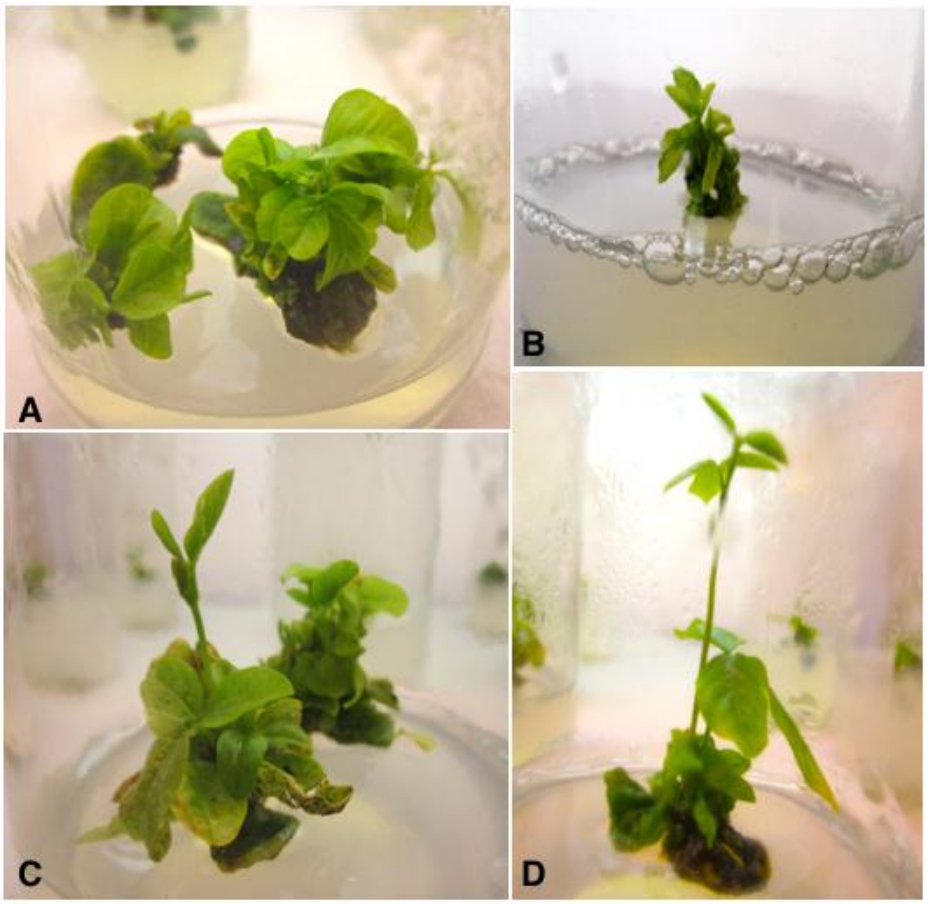


Figure 4.17: Elongation of shoots on shoot elongation medium (SEM-II) with 5 mg/l glufosinate after **(A** and **B)** 42 days, **(C)** 56 days and **(D)** 70 days of culture initiation.

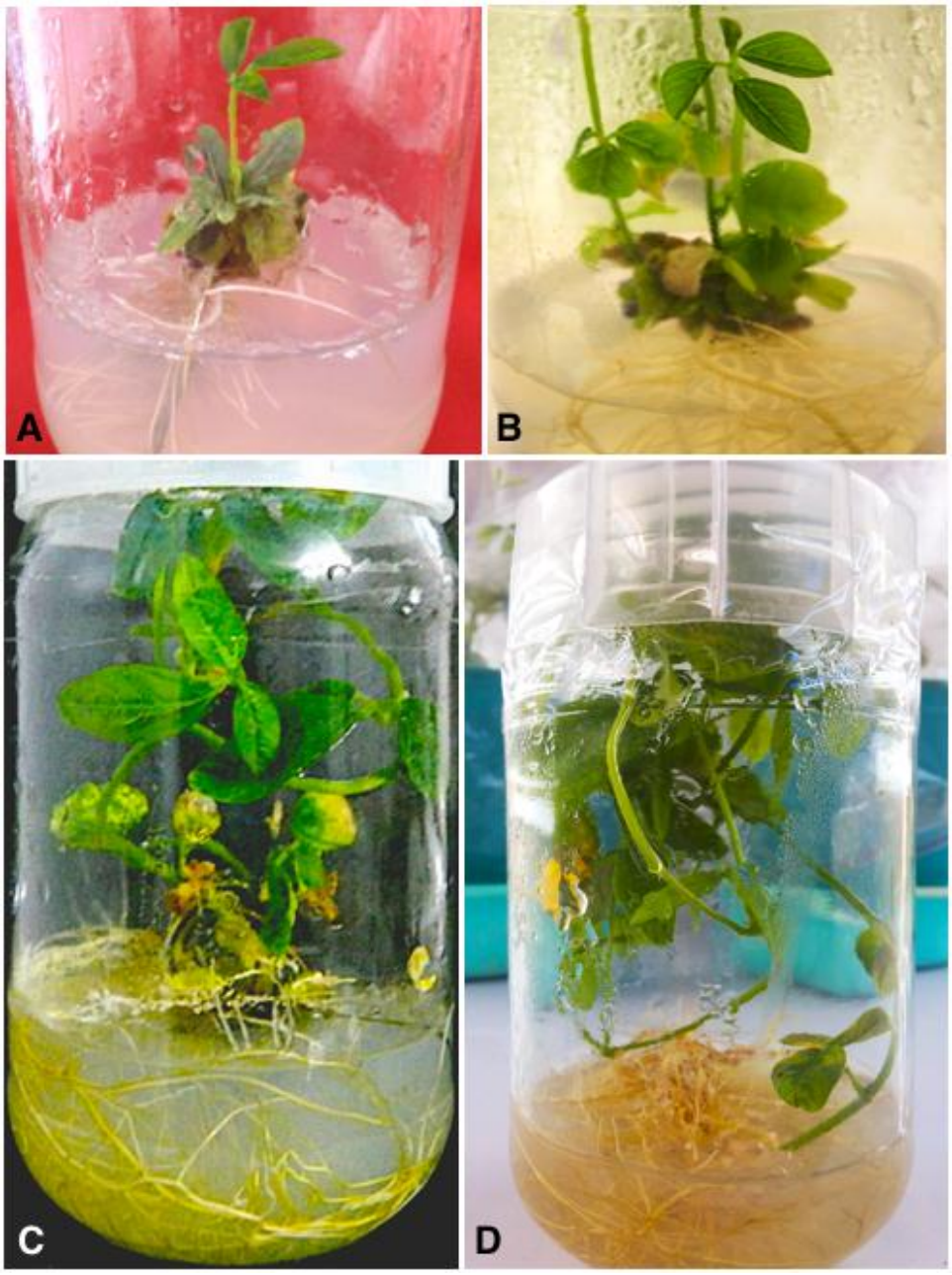


Figure 4.18: Rooting of elongated putative transgenic plantlets on RM after **(A)** 80 days, **(B)** 86 days and **(C** and **D)** 95 days of culture initiation.

Figure 4.19: Putative transgenic plants transferred to pots after 14 weeks for acclimatization and shifted to a PGW36 growth chamber under controlled conditions at National Phytotron Facility, IARI, New Delhi.

Figure 4.20: Putative T_0 transgenic plants shifted from growth chamber to glasshouse at National Phytotron Facility, IARI, New Delhi for **(A & B)** further maturation,

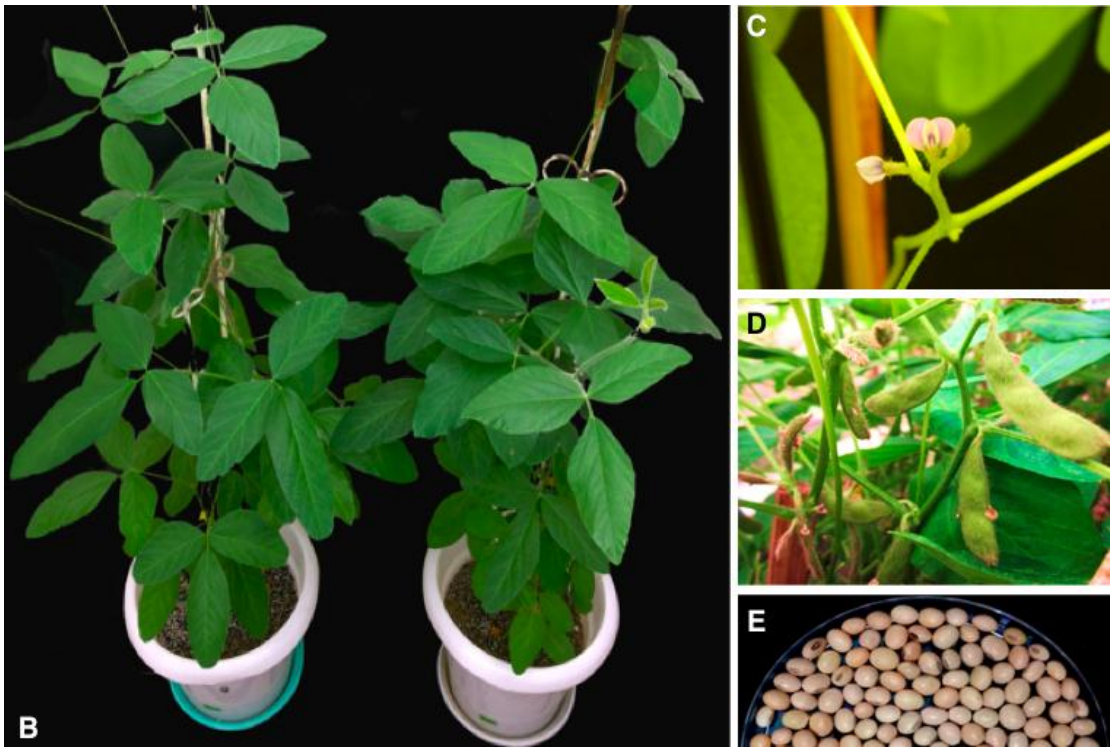


Figure 4.20: (C) flowering and **(D)** podding, to derive **(E)** fully matured T_0 soybean seeds.

A regeneration frequency of 15.5% was recorded (Table 11) as the percent of explants forming adventitious shoots compared to the total number of explants cultured [412].

Table 11: Recovery of transgenic shoots of *G. max* from cotyledonary node explants under glufosinate selection.

No. of seeds germinated (N)	No. of explants in CCM	No. of explants in SIM-I	No. of explants in SIM-II	No. of explants in SEM	No. of explants in RM	No. of plants regenerated	PCR +ve plants	Southern +ve plants
100	143	139	32	13	8	7	2	2
100	142	140	21	9	4	3	1	1
100	151	146	28	11	7	5	2	2
100	140	133	19	9	4	3	1	1
100	146	141	26	12	5	4	2	2
500	722	699	126	54	28	22	8	8

4.3.3 Transgene integration analysis

4.3.3.1 PCR examination: To check the successful integration of transgene at the

genome level, we conducted a preliminary screening by PCR amplification on the T_0 plants. We identified the putative lpa transgenic lines by amplification of ~700 bp fragment of ihp construct (Figure 4.21). The assay characterized eight independent transformation events each showing the expected amplicon size which was absent in the case of untransformed control plants.

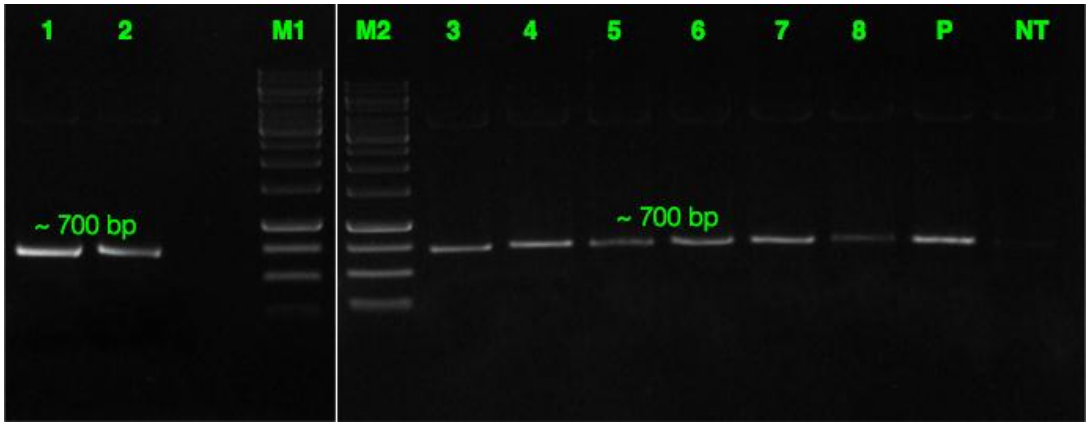


Figure 4.21: PCR amplification of *GmFad2-1 + GmIPK2_S* fragment from genomic DNA of putative T_0 transgenic plants. Lanes M1 and M2: O'GeneRuler^{TM} 1 Kb DNA ladder, lanes 1-8: positive T_0 plants showing ~700 bp *GmFad2-1 + GmIPK2_S* amplicon, P: pCWAK-IPK2ihp plasmid DNA (positive control), NT: genomic DNA from non-transformed plant (negative control).

4.3.3.2 Southern blot detection: Seeds from each independent event identified were grown successfully under containment conditions to obtain T_1 progeny plants by self-pollination (Figure 4.22). All the progenies derived appeared phenotypically normal, and fertile.

We performed segregation analysis on them by studying *bar* gene expression and found that five of the events (P2, P4, P5, P6, P8) held good fit to the stable Mendelian inheritance of a single locus (3:1) (Table 12) whilst three (P1, P3, P7) of them showed a segregation ratio of two transgene loci (15:1).

Figure 4.22: T_1 progeny plants growing in a PGW36 growth chamber under controlled conditions at National Phytotron Facility, IARI, New Delhi.

Table 12: Segregation of *bar* gene amongst T_1 progenies of eight independent transformation events characterized in *G. max* cv. Pusa 16.

Transgenic events	No. of seeds tested (n)	No. of seedlings				Ratio	Chi-square value (χ^2)*
		PCR +		PCR -			
		Obs.	Exp.	Obs.	Exp.		
P1	55	50	51.56	5	3.43	15:1	0.766
P2	47	37	35.25	10	11.75	3:1	0.347
P3	58	53	54.38	5	3.62	15:1	0.561
P4	44	35	33.00	9	11.00	3:1	0.485
P5	49	39	36.75	10	12.25	3:1	0.551
P6	52	36	39.00	16	13.00	3:1	0.923
P7	56	51	52.50	5	3.5	15:1	0.686
P8	53	37	39.75	16	13.25	3:1	0.773

*$\chi^2_{0.05;\, df1} = 3.841$.

We further confirmed the integration of the transgene cassette by southern blot analysis using a *bar* gene probe (Figure 4.23).

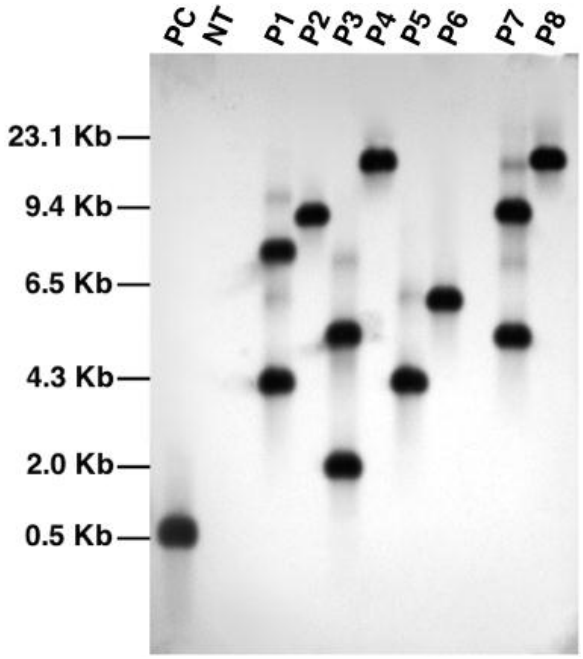


Figure 4.23: Southern blot analysis of PCR characterized transgenic events using *bar* gene-specific probe. Lane PC: plasmid pCWAK-ipk2 (positive control), lane NT: genomic DNA from non-transformed plant (negative control), lanes P1-P8: genomic DNA of independent transgenic events showing hybridization signals indicating transgene integration.

The genomic DNA was digested with *Pst*I endonuclease which is not present within the T-DNA of the construct. Upon detection, all the transgenic lines analyzed showed a distinct pattern of separation which indicates that each plant originated from an independent transformation event. We recorded the efficiency of generation of these transformation events at 1.1%, calculated as the percentage of southern confirmed transgenic plants out of the total number of explants cultured (Table 11). The blot showed that all the fragments were above 3 kb in size which suggest that the transgenic lines carry intact copies of T-DNA as the shortest fragment in each line was longer than the T-DNA region (~2.8 kb). The blot also confirmed single copy integration of the target gene into the genome of transgenic events P2, P4, P5, P6, and P8. Since single copy insertions are always desirable as they are constant over several generations of subsequent breeding [413, 414], T_2 generation plants were cultivated from seeds of these five characterized events.

4.3.4 Expression evaluation of *GmIPK2* transcript

To assess the level of reduction in *GmIPK2* expression in transgenic seeds with respect to non-transgenic control, we carried out RT-PCR and qRT-PCR analysis. To accurately quantify mRNA transcripts, it is crucial to normalize the levels of target mRNA with a stable reference gene. We carried out RT-PCR analysis by using housekeeping gene *PEPCo* as the endogenous control for normalization based on the reports by Sugimoto and co-workers [279]. We observed a variable reduction in the expression of *GmIPK2* gene with relatively constant and stable expression status of *PEPCo* transcripts for all the non-transgenic and transgenic plants which suggest that it is an ideal internal reference gene (Figure 4.24A). We, therefore, quantified the variations in *GmIPK2* transcript levels by carrying out qRT-PCR taking *PEPCo* as an

endogenous reference and found a maximum reduction of 2.79-fold and 2.56-fold in T_3 developing seeds from the transgenic lines P2-45-8 and P6-39-10 respectively (Figure 4.24B) at $P \geq 0.05$. To obtain a steady and reproducible pattern of expression, we amplified our transcript using highly specific primers which include a partial sequence of cDNA as well as its 3′ UTR.

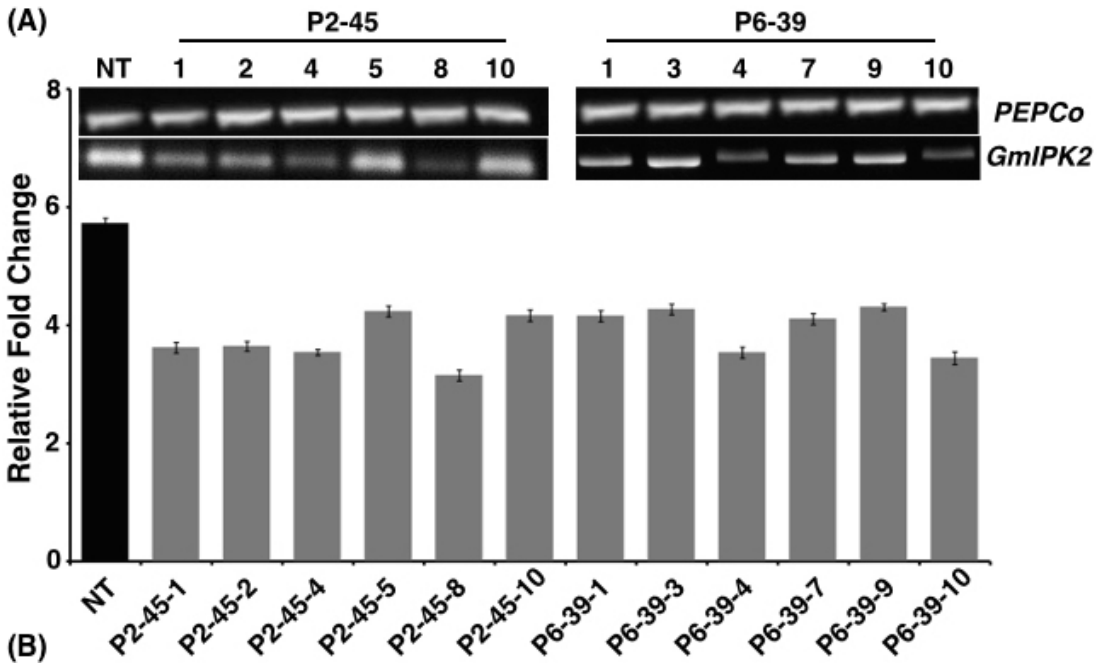


Figure 4.24: Expression analysis of transgenic soybean seeds. **(A)** RT-PCR amplification showing variation in *GmIPK2* transcripts in T_3 seeds of P2-45 and P6-39 compared to the *PEPCo* internal control (NT: Non-transformed plant). **(B)** Relative fold change measured by qRT-PCR in the samples analyzed above indicates varied levels of silencing with maximum reduction observed in P2-45-8. The data presented is mean of technical triplicates corresponding to each biological replicate (n=3) with error bars indicating standard deviation (SD).

We also studied *GmIPK2* expression in other plant tissues of these two lines and found no variation in its transcript levels (Figure 4.25). It suggests that the gene is silenced successfully only in the seeds of the transgenic plants.

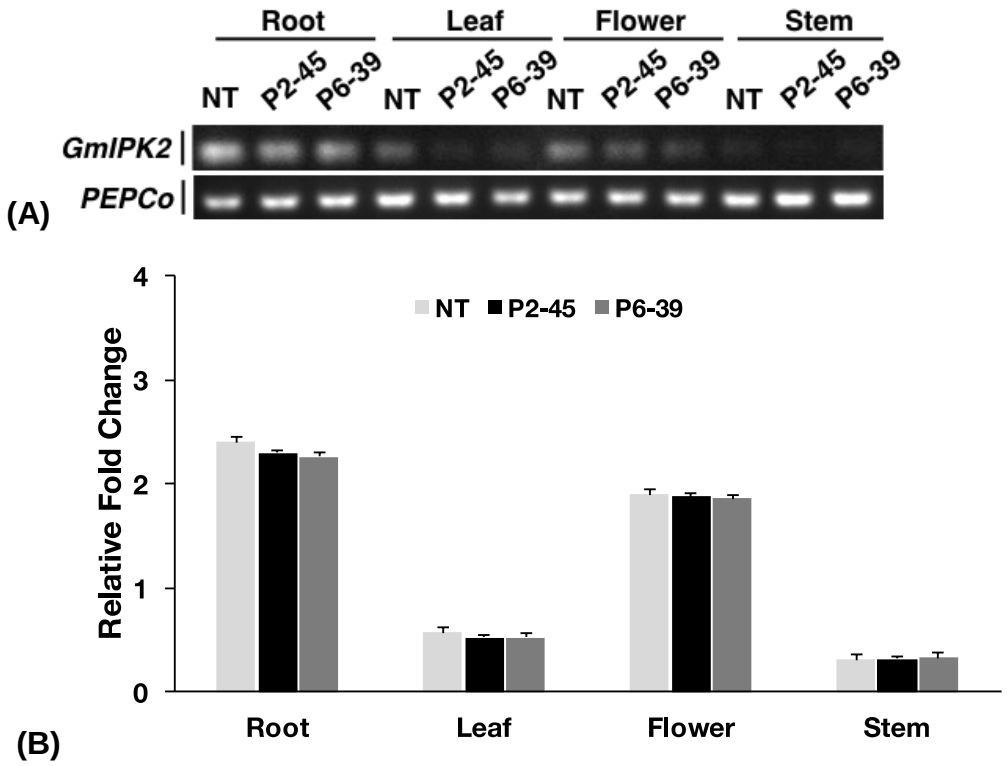


Figure 4.25: (A) RT-PCR amplification showing variation in *GmIPK2* transcripts in different tissues of transgenic events P2-45 and P6-39 compared to the *PEPCo* internal control (NT: Non-transformed plant). **(B)** Relative fold change measured by qRT-PCR in the samples analyzed above indicates no variation in the *GmIPK2* transcript level. Each sample was normalized to housekeeping gene *PEPCo*. The data presented is mean of technical triplicates corresponding to each biological replicate (n=3) with error bars indicating standard deviation (SD).

4.3.5 Quantification of seed phytic acid content

To further examine the effect of *GmIPK2* silencing on PA synthesis, we quantified PA content in mature grain extracts of above two lines and non-transgenic samples by HPLC analysis. The chromatogram obtained by scanning the samples at 197 nm displayed PA retention peaks around 7.006 ± 0.09 min, the area under which was used to compute the level of PA. We observed a decrease in the PA content in T_3 seeds compared to control which displayed a larger peak area indicative of its higher

concentration in their seeds (Figure 4.26).

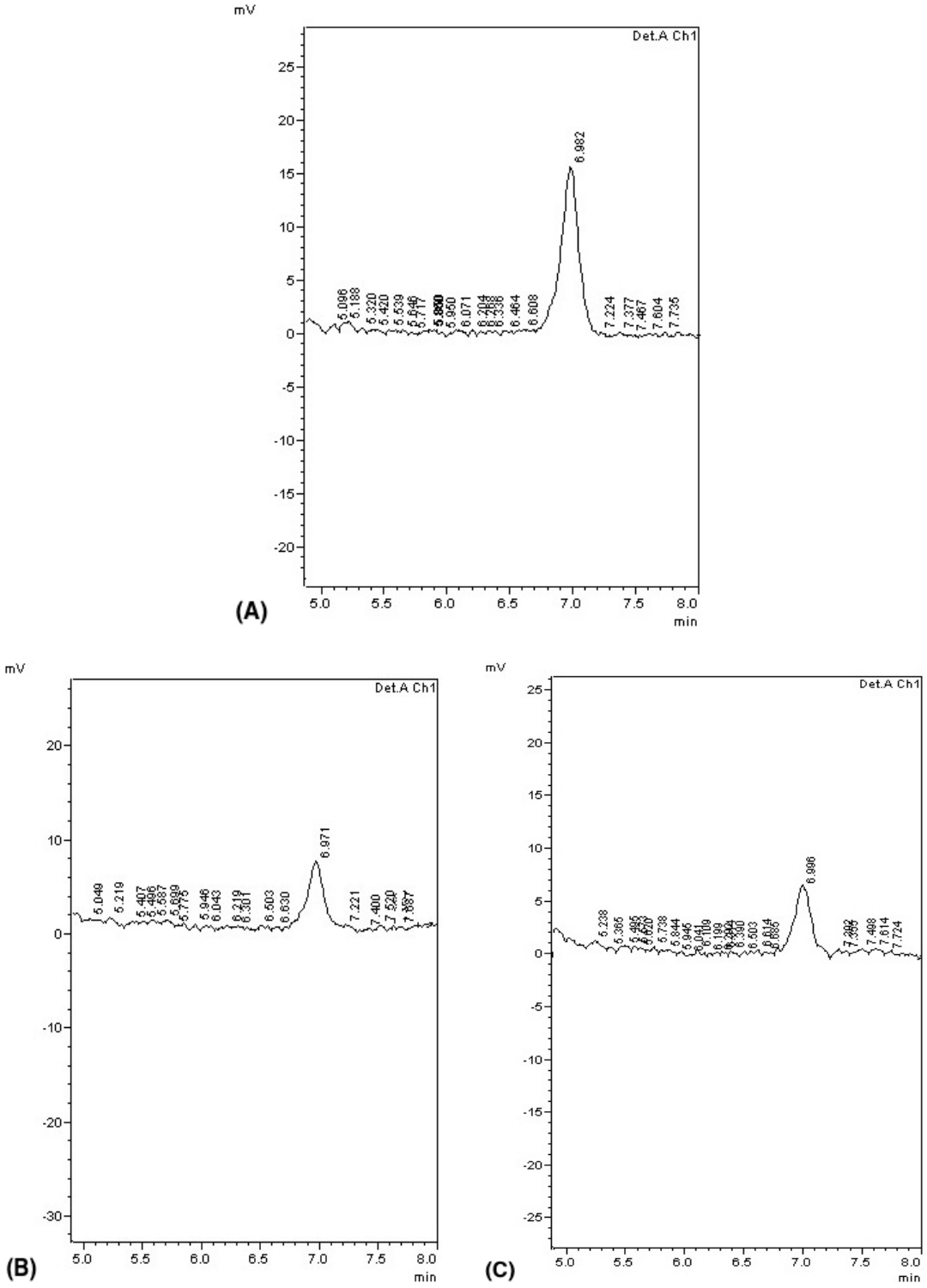


Figure 4.26: HPLC phytic acid peaks of seed extracts from **(A)** non-transgenic control and transgenic lines **(B)** P2-45-8 and **(C)** P6-39-10.

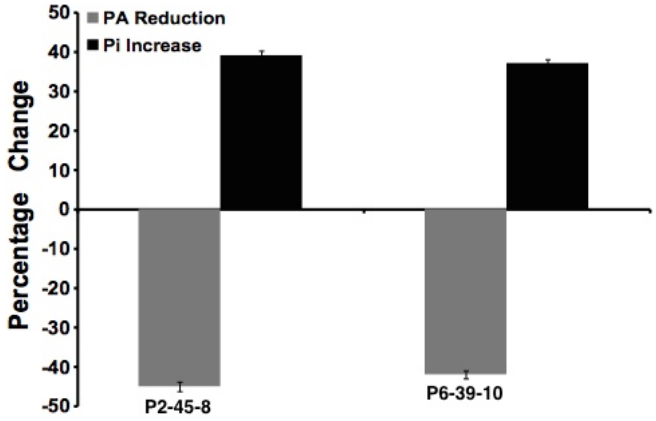


Figure 4.27: Percentage reduction in phytate (grey column) and increase in free phosphate (black column) in P2-45-8 and P6-39-10 transgenic lines.

The PA content of non-transgenic seeds was 3.48 g/100g in comparison to 1.91 g/100g and 2.02 g/100g for the seeds from transgenic lines P2-45-8 and P6-39-10 which correspond to a reduction of 45% and 42% respectively (Figure 4.27). From this, we can conclude that the decrease in the seed PA content corresponds to the decrease in *GmIPK2* expression in the seeds of these transgenic lines.

4.3.6 Estimation of seed phosphorus levels

We estimated the total P and free phosphorus (Pi) levels in transgenic as well as non-transgenic seeds to correlate it with the reduction in PA content. Estimation of seed total P is very important to the long-term goal of sustainable and environmentally friendly agricultural production.

In our study, we observed no significant difference in seed total P for the RNAi lines as compared to the control. The average total phosphorus content of T_3 seeds was 5.82 mg g-1 (P2-45-8) and 5.75 mg g-1 (P6-39-10), which was observed closely similar to that of non-transgenic seeds, 5.89 mg g-1. However, an evident change in the chemistry of lpa seeds is the increase in its Pi levels such that the total P is unchanged.

As expected, we recorded a significant increase of 39% and 37% of Pi in T_3 seeds obtained from the P2-45-8 and P6-39-10 transgenic lines respectively which gave a dark blue reaction with Chen's reagent (Figure 4.27). It thus confirms a reduction in the seed PA content of these transgenic lines.

4.3.7 Quantification of mineral content in seeds

To determine whether mitigation of PA improved the mineral bioavailability in soy grains, we made a quantitative estimation of micronutrients in transgenic seeds compared to non-transgenic control by following *in vitro* digestion method. From the observations made by atomic absorption spectrophotometer, we noticed an increase in the *in vitro* Fe, Ca and Zn availability by 16.3%, 15.05% and 8.4% for P2-45-8 and 15.9%, 14.27% and 6% for P6-39-10 seeds respectively when compared to non-transgenic control (Table 13). As mentioned before, PA binding with minerals result in the formation of insoluble salts with poor bioavailability. Our result, therefore, suggests that these lpa transgenic lines show improved mineral bioavailability.

Table 13: *In vitro* bioavailability assay of T_3 seeds.

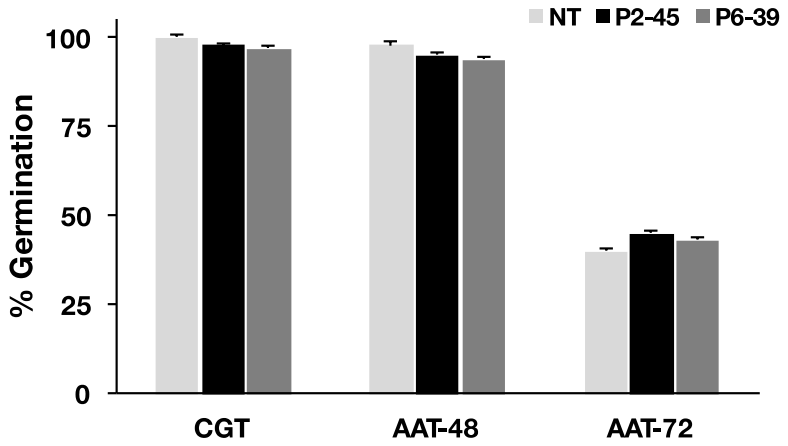
Minerals	***In vitro* availability (%)**		
	Non-transgenic (control)	**P2-45-8**	**P6-39-10**
Iron	54.80 ± 0.92	71.10 ± 1.11	70.70 ± 1.05
Calcium	44.15 ± 1.04	59.20 ± 0.78	58.42 ± 0.98
Zinc	58.50 ± 0.85	66.90 ± 0.93	64.50 ± 1.07

Data represent means ± SD, n=3.

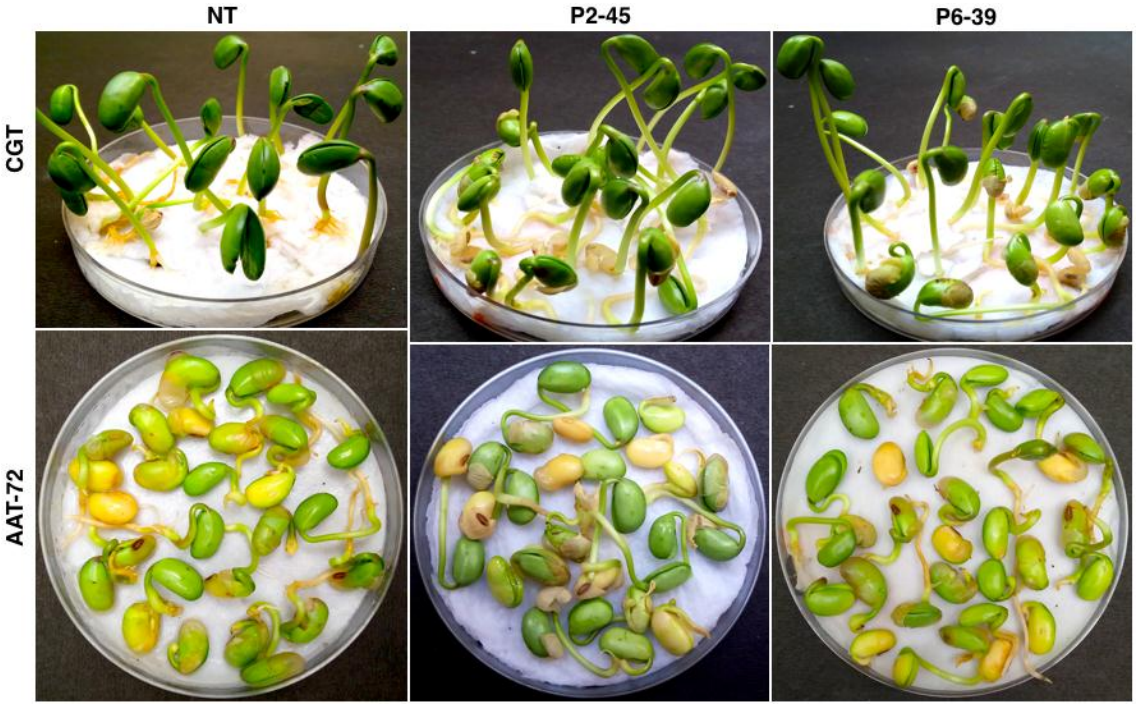
4.3.8 Seed germination and morphological trait analysis of transgenic plants

A decrease in seed PA concentration can be accompanied by adverse agronomic consequences. Therefore, plants from transgenic lines were grown in parallel with non-transgenic controls under the same environmental conditions to investigate the

extent of miscellany between them.



(A)



(B)

Figure 4.28: Analysis of germination potential in T_3 transgenic lpa seeds compared to non-transgenic seeds. **(A)** Germination percentage noticed during control germination test (CGT) and accelerated ageing test (AAT) at 48 hrs and 72 hrs time interval in both transgenic (P2-45 and P6-39) and non-transgenic seeds (NT). **(B)** Comparison of transgenic and non-transgenic seed morphology on 5th day of germination during both CGT and AAT.

We first conducted germination tests on the lpa seeds collected from our transgenics. The CGT test confirmed that 98% of the seeds show good germination while AAT test established their high vigor (Figure 4.28A). We also observed a similar germination morphology in both transgenic & control seeds for each test (Figure 4.28B).

Table 14: Morphological and yield contributing characters of T_2 transgenic plants raised in green-house.

Parameters	Non-transgenic (control)	P2-45-8	P6-39-10
Shoot length (cm)	73.40 ± 1.68	72.11 ± 1.42	70.20 ± 1.75
Root length (cm)	30.07 ± 0.15	28.43 ± 0.21	27.64 ± 0.15
Leaf length (cm)	11.20 ± 0.29	12.86 ± 0.18	12.67 ± 0.24
Leaf width (cm)	5.36 ± 0.18	6.01 ± 0.14	5.98 ± 0.17
No. of trifoliates per plant	23.07 ± 0.84	24.07 ± 0.43	23.55 ± 0.26
No. of pods per plant	49.45 ± 0.94	46.50 ± 1.08	45.18 ± 1.24
No. of seeds per pod	2.63 ± 0.08	2.72 ± 0.14	2.54 ± 0.11
100 seeds dry wt (g)	15.21 ± 0.25	14.95 ± 0.14	14.83 ± 0.32

Data represent means ± SD, n=6.

To further study the effect of PA on growth and development of seedlings, we carried out phenotyping on 75 days old plants. Interestingly, we observed morphological differences in leaves of P2 & P6 plants compared to leaves of non-transgenic plants of the same plastochron age. P2 and P6 leaves exhibited increased leaf length & leaf width and consequently leaf area, possibly due to the increase in available seed P ($P<0.05$) [415]. Leaf area, in turn, is a determinant of photosynthetic efficiency, exhibiting a direct association between the two [416]. Besides leaf size, we observed no significant differences between wild type and transgenic plants in other agronomic traits investigated ($P>0.05$) (Table 14). It confirms that *Gm-IPK2* silenced, lpa soybean seeds are capable of generating plants showing normal morphology and agronomic traits.

4.4 Conclusion

The objective of this work was to isolate successful transgenic lpa events. Since soybean is a grain-producing crop, we generated a silencing construct, pCWAK-IPK2ihp, in which transgene (*GmIPK2* ihp) is under the control of vicilin promoter to facilitate seed-specific expression. We subsequently transformed soybean cotyledonary nodes with the silencing construct following in-house standardized *Agrobacterium*-mediated gene transfer protocol and regenerated putative transgenics using the *bar* as a reporter gene. PCR, as well as southern analysis, revealed successful integration of transgene in the genome of eight transgenic events. We examined the effect of the transgene at the transcript level in these events by carrying out qRT-PCR analysis and found a maximum reduction in *GmIPK2* transcripts in developing seeds of transgenic lines P2-45-8 (2.79-fold) and P6-39-10 (2.56-fold). We, therefore, selected these lines for further investigations. To monitor the nutritional status of their seeds, we estimated PA, total P, and Pi content of the seeds by biochemical analysis. We found that the transgenic lines P2-45-8 & P6-39-10 showed a maximal reduction of 45% & 42% and a corresponding increase of 39% & 37% in their PA & Pi content respectively, while their total P level remained constant to a value observed for the non-transgenic seeds. We also noticed an increase in the *in vitro* Fe, Ca, and Zn bioavailability by 16.3%, 15.05%, and 8.4% for P2-45-8 seeds and 15.9%, 14.27%, and 6% for P6-39-10 seeds respectively. We further verified the agronomic acceptability of these lpa lines by per-forming seed germination and morphological trait analysis.

So far there is no report on successful recovery of fertile lpa soybean transgenics of Indian cultivars. Our results describe a commercially viable strategy to generate lpa grains with improved nutrient availability to address growing micro-nutrient deficiency in the world as well as to reduce its environmental impact.

This work has been published in "*Frontiers in Plant Science*", entitled "***Development and Evaluation of Low Phytic Acid Soybean by siRNA Triggered Seed Specific Silencing of Inositol Polyphosphate 6-/3-/5-Kinase Gene***".

Chapter 5

5.1 Summary and Conclusion

Soybean is a nutritionally and economically important agricultural crop. It has a unique nutrient profile and is a repository of many phytochemicals famously referred to as a virtual drugstore. Despite the multitude of nutritional benefits associated with soy consumption, the presence of infamous anti-nutritional factors presents serious challenges towards exploiting its true potential. Phytic acid, one of its most prominent anti-nutrients negatively influences the nutritional status of mono-gastrics due to the lack of phytases in their gastrointestinal tract and thus impairs their mineral bioavailability. The undigested phytate in their excreta additionally contributes significantly to the environmental phosphorus load. Therefore, developing an improved variety of soybean with reduced PA content in the seeds is imperative to augment its overall nutritional quality and overcome its associated environmental impact.

Several mutational and transgenic strategies have been investigated in the past to reduce PA in soybean. In our study, we explored RNA interference (RNAi), a reverse genetics technique for metabolic engineering of the PA pathway. This technique like others has been considered in the past to reduce PA in soybean but resulted in embryo abortion. It was accounted for by two major factors viz. the choice of gene and the promoter used for expression of the transgene. In this study, we pay focus on evaluating these two crucial factors. The biosynthesis of PA is a complex pathway controlled by various enzymes. *GmIPK2* amongst them plays a pivotal role in regulating PA turnover because of its promiscuous capability of phosphorylating multi

-ple positions on a variety of IP substrates. Therefore, we chose this gene as the target to design an intron-containing self-complementary hairpin (ihp) RNA construct to stimulate its selective degradation by post-transcriptional gene silencing & achieve a reduction in PA synthesis. Alongside, we decided to integrate a cotyledon-specific reserve protein promoter, vicilin, in the construct design to drive the target gene expression to curb phytate reduction to only seeds and prevent its constitutive effect because of which the previously reported RNAi attempts did not succeed. Thereafter, we introduced our RNAi binary construct into the soybean genome by performing *Agrobacterium tumefaciens* mediated plant transformation and identified the putative transformants by examining successful integration and expression of transgene at genome and transcript level.

At the initial stages of characterization, the plant material is a limiting factor. We thus used the PCR technique to screen putative transgenic plants at their early developmental stage by performing amplification with transgene-specific primers. We further confirmed the amplified product by conducting southern blot analysis on T_1 progeny plants grown from T_0 seeds by self-pollination under containment conditions. Southern blotting is a traditional method for determining transfer DNA (T-DNA) integration. The integration of T-DNA is a complicated mechanism and different factors including site, number and arrangement of transgene within the plant genome, the involvement of regulatory sequences, etc. contribute towards the variation in transgene expression. Therefore, studying independent transgenic lines having the same cassette of the transgene is indispensable since each one of them may act differently. Southern blotting identified single copy T-DNA insertions into the genomes of transgenic events P2, P4, P5, P6 and P8 which was further confirmed by the segregation of *bar* gene in 3:1 ratio of the monogenic pattern of

Mendelian inheritance. We thus cultivated T_2 generation plants from seeds of these five characterized events to carry out further analysis. Successful transgene integration in the plant genome, however, does not guarantee the expression of the desired traits. It is therefore pertinent to check the transgene expression and its consequences. To evaluate transgene expression in T_3 seeds as well as other plant tissues, we determined the level of reduction of its target gene (*GmIPK2*) mRNA by conventional semi-quantitative reverse-transcription polymerase chain reaction (RT-PCR) as well as dynamic quantitative real-time PCR (qRT-PCR), using steadily expressed *PEPCo* as the reference gene. The analysis revealed a variable reduction in *GmIPK2* transcript level with 2.79-fold & 2.56-fold reduction in T_3 developing seeds from the transgenic lines P2-45-8 & P6-39-10 respectively at $P \geq 0.05$. At the same time, we found no variation in *GmIPK2* transcript levels in other tissues of these two lines which suggests that the gene is successfully silenced only in the seeds, demonstrating seed-specific expression of the transgene. We further confirmed the implications of transgene expression by conducting biochemical analysis on T_3 seeds of transgenic lines P2-45-8 and P6-39-10. We expect that *GmIPK2* ihp integration/ expression will culminate in PA reduction, therefore at first, we examined the PA content in mature grain extracts of the above two lines. Their PA content was 1.91 g/100g (P2-45-8) and 2.02 g/100g (P6-39-10), which amounts to a reduction of 45% and 42% respectively when compared to non-transgenic seeds. It thus confirms a positive transgene effect. Further, research suggests that there exists a strict correspondence between seed PA and free phosphorus (Pi) [150]. Therefore, we estimated Pi levels in the above seed extracts to correlate it with the observed reduction in PA content. As expected, we recorded a significant increase of 39% (P2-45-8) and 37% (P6-39-10) in the Pi concentration. However, despite the observed

variation in Pi levels, their total phosphorus (total P) content remained unaffected, which suggests that these altered genotypes influence P partitioning in the seeds to strike a balance for supporting P-related mechanisms. Besides P, PA also acts as a sink for other minerals due to its natural strong chelation ability. Previous studies have reported improved mineral cations bioavailability of cereal lpa mutants [382-384, 417-420]. Therefore, we estimated the *in vitro* Fe, Ca and Zn availability in seeds of the above lines and noticed an increase by 16.3%, 15.05%, 8.4% (P2-45-8) and 15.9%, 14.27%, 6% (P6-39-10) respectively compared to non-transgenic control. Here we must highlight that amongst the three different minerals analyzed a maximum increase was observed in the level of Fe which support the findings of Persson and co-workers [421] who suggest that Fe in the seeds is mainly associated with PA. One of the major concerns that arise from the biochemical alterations, as observed, is the ramifications it may have on the agronomic performance of the plants. Therefore, we carried out phenotyping on 75 days old plants derived from the above lines and compared the results with the non-transgenic parent. We found no significant differences between the shoot length, root length, number of trifoliates per plant, number of pods per plant, number of seeds per pod, and 100-seed weight of the transgenic lines and the control except for increase in leaf size which is most probably the result of increase in available P content [415]. It shows that our transgene does not affect the primary metabolism of the plants and thus has a minimal effect on their agronomic performance and quality traits.

We conclude by stating that our study is the first official report of successfully achieving PA reduction in soybean by perturbing its biosynthesis pathway via RNAi to generate lines with enhanced nutritional quality and agronomic performance comparable to the commercial cultivars. Nonetheless, there is a need to conduct

extensive field testing to conclude the ultimate effect of this trait. Further studies are therefore underway in our laboratory to establish the same.

5.2 Future Research

In a follow-up study, a team of researchers can work on generating a complete set of regulated field trials data to look at all aspects of the trait, product, and crop. It is difficult but an important goal to achieve in a long-term. For efficiently screening these transgenic trials, the task of developing molecular marker linked to the target lpa loci in our transgenics can be taken up in the future since the conventional screening procedure is time-consuming and requires destructive sampling to measure the amount of phytic acid in grains. Further, we propose that the work may be extended to incorporate this trait in other grain crops having high seed PA content to achieve biofortification. Lastly, to study the impact of the block in phytic acid synthesis in seeds on the expression of genes involved in phosphorus sensing and signaling, the study of transcript profile in developing seeds of our transgenic lines should be considered as a mainstream investigation in the future.

References

1. Dwevedi, A. and Kayastha, A.M. (2011). Soybean: a multifaceted legume with enormous economic capabilities. In: Tzi-Bun Ng. ed. *Soybean: Biochemistry, Chemistry and Physiology*. InTechOpen, Croatia. pp. 165-188.

2. Sebastian, S., P.S. Kerr, R.W. Pearlstein, and W.D. Hitz. (2000). Soybean germplasm with novel genes for improved digestibility. In: Drackley J.K. ed. *Soy in Animal Nutrition*. Federation of Animal Science Societies, Savoy, IL. pp. 56-74.

3. Heaney, R.P., Weaver, C.M. and Fitzsimmons, M.L. (1991). Soybean phytate content effect on calcium absorption. *Am J Clin Nutr*. 53(3):745-747.

4. Lynch, S.R., Dassenko, S.A., Cook, J.D., Juillerat, M.A. and Hurrell, R.F. (1994). Inhibitory effect of a soybean protein related moiety on iron-absorption in humans. *Am J Clin Nutr*. 60(4):567-572.

5. Weaver, C.M. and Kannan, S. (2002). Phytate and mineral bioavailability. In: Reddy N.R., Sathe S.K. eds. *Food Phytates*. CRC Press, Boca Raton, FL. pp. 211-223.

6. Hurrell, R.F. (2003). Influence of vegetable protein sources on trace element and mineral bioavailability. *J Nutr*. 133(9):2973S-2977S.

7. Yan, F., Fritts, C.A. and Waldroup, P.W. (2003). Evaluation of modified dietary phosphorus levels with and without *phytase* supplementation on live performance and fecal phosphorus levels in broiler diets. 1. Full-term feeding recommendations. *J Appl Poultry Res*. 12:174-182.

8. Maguire, R.O., Crouse, D.A. and Hodge., S.C. (2007). Diet modification to reduce phosphorus surpluses: A mass balance approach. *J Environ Qual*. 36:1235-1240.

9. Ertl, D.S., Young, K.A. and Raboy, V. (1998). Plant genetic approaches to phosphorus management in agricultural production. *J Environ Qual*. 27:299-304.

10. Sharpley, A.N., Daniel, T., Sims, T., Lemunyon, J., Stevens, R. and Perry, R. (2003). *Agricultural Phosphorus and Eutrophication 2nd Edition*. USDA-ARS publication ARS-149.

11. Hotz, C. and Gibson, R.S. (2007). Traditional food-processing and preparation practices to enhance the bioavailability of micronutrients in plant-based diets. *J Nutr*. 137:1097-1100.

12. Mendoza, C., F.E. Viteri, B. Lonnerdal, V. Raboy, K.A. Young, and K.H. Brown. (2001). Absorption of iron from unmodified maize and genetically altered low-phytate maize fortified with ferrous sulfate of sodium iron EDTA. *Am. J. Clin. Nutr.* 73:80-85.

13. Baxter, C.A., Joern, B.C., Ragland, D., Sands, J.S. and Adeola, O. (2003). *Phytase*, high-available phosphorus corn, and storage effects on phosphorus levels in pig excreta. *J Environ Qual*. 32:1481-1489.

14. Dungelhoef, M., Rodehutscord, M., Spiekers, H. and Pfeffer, E. (1994). Effects of supplemental microbial phytase on availability of phosphorus contained in maize, wheat and triticale to pigs. *Anim Feed Sci Tech*. 49:1-10.

15. Yi, Z., Kornegay, E.T., Ravindran, V. and Denbow, D.M. (1996). Improving phytate phosphorus availability in corn and soybean meal for broilers using microbial *phytase* and calculation of phosphorus equivalency values for *phytase*. *Poultry Sci*. 75:240-249.

16. Denbow, D.M., Grabau, E.A., Lacy, G.H., Kornegay, E.T., Russell, D.R. and Umbeck, P. (1998). Soybeans transformed with a fungal *phytase* gene improve phosphorus availability for broilers. *Poultry Sci.* 77:878-881.

17. Kornegay, E.T. (2001). Digestion of phosphorus and other nutrients: the role of *phytases* and factors influencing their activity. In: Bedford M.R., Partridge G.G. eds. *Enzymes in farm animal nutrition*. CABI Publishing. pp 237-272.

18. Boyce, A. and Walsh, G. (2006). Comparison of selected physicochemical characteristics of commercial *phytases* relevant to their application in phosphate pollution abatement. *J Environ Sci Health A Tox Hazard Subst Environ Eng*. 41:789-798.

19. Raboy, V. (2009). Approaches and challenges to engineering seed phytate and total phosphorus. *Plant Sci*. 177:281-296.

20. Golovan, S.P., Meidinger, R.G., Ajakaiye, A., Cottrill, M., Wiederkehr, M.Z., Barney, D.J., Plante, C., Pollard, J.W., Fan, M.Z., Hayes, M.A., Laursen, J., Hjorth, J.P., Hacker, R.R., Phillips, J.R. and Forsberg, C.W. (2001). Pigs expressing salivary *phytase* produce low-phosphorus manure. *Nat Biotechnol*. 19:741-745.

21. Hegeman, C.E., Good, L.L. and Grabau, E.A. (2001). Expression of *D-myo-inositol-3-phosphate synthase* in soybean: Implications for phytic acid biosynthesis. *Plant Physiol*. 125:1941-1948.

22. Bilyeu, K.D., Zeng, P., Coello, P., Zangh, Z.J., Krishnan, H.B., Beuselinck, P. R. and Polacco, J.C. (2008). Quantitative conversion of phytate to utilizable phosphorus in soybean seeds expressing a bacterial *phytase*. *Plant Physiol*. 146:468-477.

23. Nunes, A., Vianna, G., Cuneo, F., Amaya-Farfan, J., de Capdeville, G., Rech, E. and Aragao, F. (2006). RNAi-mediated silencing of the *myo-inositol-1-phosphate synthase* gene (*GmMIPS1*) in transgenic soybean inhibited seed development and reduced phytate content. *Planta*. 224:125-132.

24. Kuwano, M., Mimura, T., Takaiwa, F. and Yoshida, K. (2009). Generation of stable "low phytic acid" transgenic rice through antisense repression of the *1D-myo-inositol 3-phosphate synthase* gene (*RINO1*) using the 18-kda oleosin promoter. *Plant Biotechnol J*. 7:96-105.

25. Shi, J., Wang, H., Schellin, K., Li, B., Faller, M., Stoop, J.M., Meeley, R.B., Ertl, D.S., Ranch, J.P. and Glassman, K. (2007). Embryo-specific silencing of a transporter reduces phytic acid content of maize and soybean seeds. *Nat Biotechnol*. 25:930-937.

26. George, T.S., Richardson, A.E., Li, S.S., Gregory, P.J. and Daniell, T.J. (2009). Extracellular release of a heterologous *phytase* from roots of transgenic plants: Does manipulation of rhizosphere biochemistry impact microbial community structure? *FEMS Microbiol Ecol*. 70:433-445.

27. George, T.S., Simpson, R.J., Hadobas, P.A. and Richardson, A.E. (2005). Expression of a fungal *phytase* gene in *Nicotiana tabacum* improves phosphorus nutrition of plants grown in amended soils. *Plant Biotechnol J*. 3:129-140.

28. Holme, I.B., Dionisio, G., Brinch-Pedersen, H., Wendt, T., Madsen, C.K., Vincze, E. and Holm, P.B. (2012). Cisgenic barley with improved *phytase* activity. *Plant Biotechnol J*. 10(2):237-47.

29. Tang, G., Galili, G. and Zhuang, X. (2007). RNAi and microRNA: breakthrough technologies for the improvement of plant nutritional value and metabolic engineering. *Metabolomics*. 3:357-369.

30. Vaucheret, H., Beclin, C. and Fagard, M. (2001). Post-transcriptional gene silencing in plants. *J Cell Sci*. 114:3083-3091.

31. Smith, N. A., Singh, S. P., Wang, M. B., Stoutjesdijk, P. A., Green, A. G. and Waterhouse, P. M. (2000). Total silencing by intron spliced hairpin RNAs. *Nature*. 407:319-320.

32. Wesley, S.V., Liu, Q., Wielopolska, A., Ellacott, G., Smith, N., Singh, S. and Helliwell, C. (2003). Custom knock-outs with hairpin RNA-mediated gene silencing. *Methods Mol Biol*. 236:273-286.

33. Chen, S., Hofius, D., Sonnewald, U. and Bornke, F. (2003). Temporal and spatial control of gene silencing in transgenic plants by inducible expression of double-stranded RNA. *Plant J*. 36:731-740.

34. Guo, H.S., Fei, J.F., Xie, Q. and Chua, N.H. (2003). A chemical regulated inducible RNAi system in plants. *Plant J*. 34:383-392.

35. Wielopolska, A., Townley, H., Moore, I., Waterhouse, P. and Helliwell, C. (2005). A high-throughput inducible RNAi vector for plants. *Plant Biotechnol J*. 3:583-590.

36. Prattley, C.A. and Stanley, D.W. (1982). Protein-phytate interactions in soybeans. I. Localization of phytate in protein bodies and globoids. *J Food Biochem*. 6:243-253.

37. Arumuganathan, K. and Earle, E.D. (1991). Nuclear DNA content of some important plant species. *Plant Mol Biol Report*. 9:208-218.

38. Hartig, T. (1855). Uber das klebermehl. *Bot Ztg*. 13:881.

39. Cosgrove, D.J. (1980). Inositol hexakisphosphates. In: Cosgrove D.J. ed. *Inositol Phosphates: Their Chemistry, Biochemistry and Physiology*. Elsevier Scientific Publishing Company, Amsterdam. pp. 26-43.

40. Raboy, V. (1997). Accumulation and storage of phosphate and minerals. In: Larkins B.A., Vasil I.K. eds. *Cellular and Molecular Biology of Plant Seed Development*. Kluwer Academic Publishers, Dordrecht, Netherlands. pp. 441-477.

41. O'Dell, B.L., de Boland, A.R. and Koirtyohann, S.R. (1972). Distribution of phytate and nutritionally important elements among the morphological components of cereal grains. *J Agric Food Chem*. 20:718-721.

42. Lott, J.N.A. (1984). Accumulation of seed reserves of phosphorus and other minerals. In: Murray D.R. ed. *Seed Physiology*, Volume1. Academic Press, Sydney, Australia. pp. 154-196.

43. Shi, J., Wang, H., Wu, Y., Hazebroek, J., Meeley, R.B. and Ertl, D.S. (2003). The maize low-phytic acid mutant lpa2 is caused by mutation in an inositol phosphate kinase gene. *Plant Physiol*. 131:507-515.

44. Campbell M., Dunn R., Ditterline R., Pickett S. and Raboy V. (1991). Phytic acid represents 10 to 15% of total phosphorus in alfalfa root and crown. *J Plant Nutr*. 14:925-937.

45. Bentsink L., Yuan K., Koornneef M. and Vreugdenhil D. (2003). The genetics of phytate and phosphate accumulation in seeds and leaves of *Arabidopsis thaliana*, using natural variation. *Theor Appl Genet*. 106:1234-1243.

46. Samotus B. and Schwimmer S. (1962). Effect of maturity and storage on distribution of phosphorus among starch and other components of potato tuber. *Plant Physiol*. 37:519-522.

47. Reddy N.R., Sathe S.K. and Salunkhe D.K. (1982). *Phytases* in legumes and cereals. *Adv Food Res*. 82:1-92.

48. Angel, R., Tamim, N.M., Applegate, T.J., Dhandu, A.S. and Ellestad, L.E. (2002). Phytic acid chemistry: influence on phytin-phosphorus availability and *phytase* efficacy. *J Appl Poult Res*. 11(4):471-480.

49. Ananieva, E.A. and Gillaspy, G.E. (2009). Switches in nutrient and inositol signaling. *Plant Signal Behav*. 4:304-306.

50. Lemtiri-Chlieh, F., MacRobbie, E., Webb, A., Manison, N., Brownlee, C., S-kepper, J., Chen, J., Prestwich, G. and Brearley, C. (2003). Inositol hexakisphosphate mobilizes an endo-membrane store of calcium in guard cells. *Proc Natl Acad Sci*. 100(17):10091-10095.

51. Thole, J. and Nielsen, E. (2008). Phosphoinositides in plants: Novel functions in membrane trafficking. *Curr Opin Plant Biol*. 11:620-631.

52. Murphy, A., Otto, B., Brearley, C., Carr, J. and Hanke, D. (2008). A role for inositol hexa*kis*phosphate in the maintenance of basal resistance to plant pathogens. *Plant J*. 56:638-652.

53. Doria, E., Galleschi, L., Galucci, L., Pinzino, C., Pilu, R., Cassani, E. and Nielsen, E. (2009). Phytic acid prevents oxidative stress in seeds: evidence from a maize (Zea mays L.) low phytic acid mutant. *J Exp Bot*. 60:967-978.

54. Munnik, T. and Vermeer, J. (2010). Osmotic stress-induced phosphoinositide and inositol phosphate signalling in plants. *Plant Cell Environ*. 33:655-669.

55. Kuo, H., Chang, T., Chiang, S., Wang, W., Charng, Y. and Chiou, T. (2014). Arabidopsis *inositol pentakisphosphate 2-kinase*, *AtIPK1*, is required for growth and modulates phosphate homeostasis at the transcriptional level. *Plant J*. 80:503-515.

56. Qin, Z.X., Chen, Q.J., Tong, Z. and Wang, X.C. (2005). The Arabidopsis *inositol 1,3,4-trisphosphate 5/6 kinase*, *AtITPK-1*, is involved in plant photo-morphogenesis under red light conditions, possibly via interaction with COP9 signalosome. *Plant Physiol Biochem*. 43:947-954.

57. Lee, H., Lee, D., Cho, H., Kim, S., Auh, J. and Pai, H. (2015). InsP6-sensitive variants of the Gle1 mRNA export factor rescue growth and fertility defects of the ipk1 low-phytic-acid mutation in *Arabidopsis*. *Plant Cell*. 27, 417-431.

58. Shears, S. (2001). Assessing the omnipotence of inositol hexakisphosphate. *Cell Signal*. 13: 151-158.

59. Schröterová, L., Hasková, P., Rudolf, E. and Cervinka, M. (2010). Effect of phytic acid and inositol on the proliferation and apoptosis of cells derived from colorectal carcinoma. *Oncol Rep*. 23:787-793.

60. Verghese, M., Rao, D.R., Chawan, C.B., Walker, L.T. and Shackelford, L. (2006). Anticarcinogenic effect of phytic acid (IP6): Apoptosis as a possible mechanism of action. *LWT*. 39:1093-1098.

61. Xu, Q., Kanthasamy, A.G. and Reddy, M.B. (2008). Neuroprotective effect of the natural iron chelator, phytic acid in a cell culture model of Parkinson's disease. *Toxicology*. 245:101-108.

62. Macbeth, M.R., Schubert, H.L., VanDemark, A.P., Lingam, A.T., Hill, C.P. and Bass, B.L. (2005). Inositol hexa*kis*phosphate is bound in the ADAR2 core and required for RNA editing. *Science*. 309:1534-1539.

63. York, J.D., Odom, A.R., Murphy, R., Ives, E.B. and Wente, S.R. (1999). A phospholipase C-dependent inositol polyphosphate kinase pathway required for efficient messenger RNA export. *Science*. 285:96-100.

64. Hanakahi, L.A., Bartlet-Jones, M., Chappell, C., Pappin, D. and West, S.C. (2000). Binding of inositol phosphate to DNA-PK and stimulation of double-strand break repair. *Cell*. 102:721-729.

65. Shen, X., Xiao, H., Ranallo, R., Wu, W.H. and Wu, C. (2003). Modulation of ATP-dependent chromatin-remodeling complexes by inositol polyphosphates. *Science*. 299:112-114.

66. Steger, D.J., Haswell, E.S., Miller, A.L., Wente, S.R. and O'Shea, E.K. (2003). Regulation of chromatin remodeling by inositol polyphosphates. *Science*. 299:114-116.

67. Vucenik, I. and Shamsuddin, A.M. (2006). Protection against cancer by dietary IP6 and inositol. *Nutr Canc*. 55:109-125.

68. Messina, M.J. (1995). Soy foods: their role in disease prevention and treatment. In: Liu K. ed. *Soybeans: Chemistry, Technology, and Utilisation*. Chapman and Hall, New York. pp. 442-447.

69. Erdman, J.W. (1979). Oilseed *phytates*: Nutritional implications. *J Am Oil Chem Soc*. 56:736-741.

70. Maddaiah, V.T., Kurnick A.A., Hulett B.J. and Reid B.L. (1964). Nature of intestinal phytase activity. *Proc Soc Exp Biol Med*. 115:1054-1057.

71. Vohra, P., Gray, A. and Kratzer, F.H. (1965). Phytic acid metal complexes. *Proc Soc Exp Biol Med*. 120:447.

72. Wise, A. (1983). Dietary factors determining the biological activities of *phytase*. *Nutr Abstr Rev*. 53:791-806.

73. Simpson, C.J. and Wise, A. (1990). Binding of zinc and calcium to inositol phosphates (phytate) in vitro. *Br J Nutr*. 64:225-232.

74. Persson, H., Turk, M., Nyman, M. and Sandberg, A.S. (1998). Binding of Cu 2^{+}, $Zn2^{+}$, and $Cd2^{+}$ to inositol tri-, tetra-, penta-, and hexaphosphate. *J Agric Food Chem*. 46:3194-3200.

75. Maenz, D.D., Engle-Schaan, C.M., Newkirk, R.W. and Classen, H.L. (1999). The effect of minerals and mineral chelators on the formation of *phytase*-resistant and *phytase* susceptible forms of phytic acid in solution and in a slurry of canola meal. *Anim Feed Sci Technol*. 81:177-192.

76. Davies, N.T. and Flett, A.A. (1978). The similarity between *alkaline phosphatase* (EC.3.1.3.1) and *phytase* (EC.3.1.3.8) activity in rat intestine and their importance in phytate induced zinc deficiency. *Br J Nutr*. 39:307-316.

77. Rehman, A.U., Shunmugam, A., Arganosa, G., Bett, K.E. and Warkentin, T.D. (2012). Inheritance of the low-phytate trait in pea. *Crop Sci*. 52:1171-1175.

78. Ford, J.A., Colhoun, E.M., McIntosh, W.B., and Dunnigan, M.G. (1972). Rickets and osteomalacia in the Glasgow Pakistani community, 1961-71. *Br Med J*. 3:677-680.

79. Wills, M.R., Day, R.C., Phillips, J.B., and Bateman, E.C. (1972). Phytic acid and nutritional rickets in immigrants. *Lancet*. 1(7754):771-3.

80. Reinhold, J.G., Nasr, K., Lahimgarzadeh, A. and Hedayati, H. (1973). Effects of purified phytate and phytate-rich bread upon metabolism of zinc, calcium, phosphorus, and nitrogen in man. *Lancet*. 1(7798):283-8.

81. Martin, C.J. and Evans, W.J. (1989). Phytic acid-enhanced metal ion exchange reactions: the effect on *carboxypeptidase A1*. *J Inorg Biochem*. 35:267-288.

82. Vohra, A. and Satyanarayan, T. (2003). *Phytases*: microbial sources, production, purification and potential biotechnological applications. *Crit Rev Biotechnol*. 23:29-60.

83. Lott, J.N.A., Ockenden, I., Raboy, V. and Batten, G.D. (2002). Phytic acid and phosphorus in crop grains, seeds and fruits. In: Reddy N.R., Sathe S.K. eds. *Food Phytate*. CRC Press, Boca Raton, FL. pp. 7-24.

84. Daniel, T.C., Sharpley, A.N. and Lemunyo, J.L. (1998). Agricultural phosphorus and eutrophication: A symposium overview. *J Environ Qual*. 27:251-257.

85. Majumder, A.N.L., Mandal, N.C. and Biswas, B.B. (1972). *Phosphoinositol kinase* from germinating mung bean seeds. *Phytochemistry*. 11:503-508.

86. Brearley, C.A. and Hanke, D.E. (1996). Metabolic evidence for the order of addition of individual phosphate esters to the *myo*-inositol moiety of inositol hexa*kis*phosphate in the duckweed *Spirodela polyrhiza* L. *Biochem J*. 314: 227-233.

87. Stevenson-Paulik, J., Bastidas, R., Chiou, S., Frye, R. and York, J. (2005). Generation of phytate-free seeds in *Arabidopsis* through disruption of inositol polyphosphate kinases. *Proc Natl Acad Sci*. 102:12612-12617.

88. Kim, S. and 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.

89. Suzuki, M., Tanaka, K., Kuwano, M. and Yoshida, K. (2007). Expression pattern of inositol phosphate-related enzymes in rice (*Oryza sartiva* L.): Implications for the phytic acid biosynthetic pathway. *Gene*. 405:55-64.

90. Stiles, A., Qian, X., Shears, S. and Grabau, E. (2008). Metabolic and signaling properties of an *ITPK* gene family in *Glycine max*. *FEBS Lett*. 582:1853-1858.

91. Fileppi, M., Galasso, I., Tagliabue, G., Daminati, M., Campion, B., Doria, E. and Sparvoli, F. (2010). Characterisation of structural genes involved in phytic acid biosynthesis in common bean (*Phaseolus vulgaris* L.). *Mol Breed*. 25: 453-470.

92. Bhati, K., Aggarwal, S., Sharma, S., Mantri, S., Singh, S., Bhalla, S., Kaur, J., Tiwari, S., Roy, J., Tuli, R., et al. (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.

93. Stephens, L.R. and Irvine, R.F. (1990). Stepwise phosphorylation of myo-inositol leading to *myo*-inositol hexa*kis*phosphate in *Dictyostelium*. *Nature*. 346:580-583.

94. Odom, A.R., Stahlberg, A., Wente, S.R. and York, J.D. (2000). A role for nuclear *inositol 1,4,5-trisphosphate kinase* in transcriptional control. *Science.* 287:2026-2029.

95. Seeds, A.M., Sandquist, J.C., Spana, E.P. and York, J.D. (2004). A molecular basis for inositol polyphosphate synthesis in *Drosophila melanogaster*. *J Biol Chem*. 279:47222-47232.

96. Verbsky, J.W., Chang, S.C., Wilson, M.P., Mochizuki, Y. and Majerus, P.W. (2005). The pathway for the production of inositol hexa*kis*phosphate in human cells. *J Biol Chem*. 280: 1911-1920.

97. Loewus, F.A. and Murthy, P.P.N. (2000). *Myo*-inositol metabolism in plants. *Plant Sci*. 150:1-19.

98. Raboy, V. (2001). Seeds for a better future: 'low phytate' grains help to overcome malnutrition and reduce pollution. *Trends Plant Sci*. 6:458-462.

99. Raboy, V. (2002). Progress in breeding low phytate crops. *J Nutr*. 132:503S-505S.

100. Hitz, W., Carlson, T.J., Kerr, P.S. and Sebastian, S.A. (2002). Biochemical and molecular characterization of a mutation that confers a decreased raffino-saccharide and phytic acid phenotype on soybean seeds. *Plant Physiol.* 128: 650-660.

101. Larson, S. and Raboy, V. (1999). Linkage mapping of maize and barley *myo-inositol 1-phosphate synthase* DNA sequences: Correspondence with a low phytic acid mutation. *Theor Appl Genet*. 99:27-36.

102. Yoshida, K., Wada, T., Koyama, H., Mizobuchi-Fukuoka, R. and Naito, S. (1999). Temporal and spatial patterns of accumulation of the transcript of *myo-inositol-1-phosphate synthase* and phytin-containing particles during seed development in rice. *Plant Physiol*. 119:65-72.

103. Chappell, A., Scaboo, A., Wu, X., Nguyen, H., Pantalone, V. and Bilyeu, K. (2006). Characterization of the MIPS gene family in *Glycine max*. *Plant Breed*. 125:493-500.

104. Luo, Y., Qin, G., Zhang, J., Liang, Y., Song, Y., Zhao, M., Tsuge, T., Aoyama, T., Liu, J., Gu, H., et al. (2011). D-myo-inositol-3-phosphate affects phosphatidylinositol-mediated endomembrane function in *Arabidopsis* and is essential for auxin-regulated embryogenesis. *Plant Cell*. 23:1352-1372.

105. Sparvoli, F. and Cominelli, E. (2015). Seed biofortification and phytic acid reduction: a conflict of interest for the plant? *Plants*. 4:728-755.

106. Yuan, F., Zhu, D., Tan, Y., Dong, D., Fu, X., Zhu, S., Li, B. and Shu, Q. (2012). Identification and characterization of the soybean IPK1 ortholog of a low phytic acid mutant reveals an exon-excluding splice-site mutation. *Theor Appl Genet*. 125:1413-1423.

107. Xu, X.H., Zhao, H.J., Liu, Q.L., Frank, T., Engel, K.H., An, G.H. and Shu, Q. Y. (2009). Mutations of the multi-drug resistance-associated protein ABC transporter gene 5 result in reduction of phytic acid in rice seeds. *Theor Appl Genet*. 119:75-83.

108. Stevenson-Paulik, J., Odom, A.R. and York, J.D. (2002). Molecular and biochemical characterization of two plant *inositol polyphosphate 6-/3-/5-kinases*. *J Biol Chem*. 277:42711-42718.

109. Saiardi, A., Caffrey, J.J., Snyder, S.H. and Shears, S.B. (2000). *Inositol polyphosphate multikinase* (ArgRIII) determines nuclear mRNA export in *Saccharomyces cerevisiae*. *FEBS Letters*. 468:28-32.

110. Takazawa, K., Perret, J., Dumont, J.E. and Erneux, C. (1991). Molecular cloning and expression of a new putative *inositol 1,4,5-trisphosphate 3-kinase* isoenzyme. *Biochem J.* 278:883-886.

111. Dewaste, V., Pouillon, V., Moreau, C., Shears, S., Takazawa, K. and Erneux, C. (2000). Cloning and expression of a cDNA encoding human inositol *1,4,5-trisphosphate 3-kinase C*. *Biochem J.* 352:343-351.

112. Choi, K.Y., Kim, H.K., Lee, S.Y., Moon, K.H., Sim, S.S., Kim, J.W., Chung, H.K. and Rhee, S.G. (1990). Molecular cloning and expression of a complementary DNA for *inositol 1,4,5-trisphosphate 3-kinase*. *Science.* 248:64-66.

113. Thomas, S., Brake, B., Luzio, J.P., Stanley, K. and Banting, G. (1994). Isolation and sequence of a full-length cDNA encoding a novel rat *inositol 1,4,5-trisphosphate 3-kinase*. *Biochim Biophys Acta.* 1220:219-222.

114. Bertsch, U., Haefs, M., Moller, M., Deschermeier, C., Fanick, W., Kitzerow, A., Ozaki, S., Meyer, H.E. and Mayr, G.W. (1999). A novel A-isoform-like *inositol 1,4,5-trisphosphate 3-kinase* from chicken erythrocytes exhibits alternative splicing and conservation of intron positions between vertebrates and invertebrates. *Gene.* 228:61-71.

115. Haard, N.F., Odunfa, S.A., Lee, C.H., Quintero-Ramirez, R., Lorence-Quinones, A. and Wacher-Radarte, C. (1989). Fermented cereals: a global perspective. *FAO Agricultural Service Bulletin,* Issue 138. FAO, Rome, Italy. pp. 63-97.

116. Makokha, A.O., Oniango, R.K., Njoroge, S.M. and Kamar, O.K. (2002). Effect of traditional fermentation and malting on phytic acid and mineral availability from sorghum (*Sorghum bicolor*) and funger millet (*Eleusine caracana*) grain varieties grown in Kenya. *Food Nutr Bull.* 23:241-245.

117. Coulibaly, A., Kouakou, B. and Chen J. (2011). Phytic acid in cereal grains: Healthy or harmful ways to reduce phytic acid in cereal grains and their effects on nutritional quality. *Am J plant Nutr Fert Technol*. 1:1-22.

118. Marshall, A.A., Samuel, J.E., Mary, U.E. and Inegbenose, G.I. (2011). Effect of germination on the phytase activity, phytate and total phosphorus contents of rice, maize, millet, sorghum and wheat. *J Food Sci Tech.* 48:724-729.

119. Nout, M.J.R. (1993). Processed weaning foods for tropical climates. *Int J Food Sci Nutr*. 43:213-221.

120. Mahgoub, S.E.O. and Elhag, S.A. (1998). Effect of milling, soaking, malting, heat-treatment and fermentation on phytate level of four Sudanese sorghum cultivars. *Food Chem.* 61:77-80.

121. Perlas, L.A. and Gibson, R.S. (2002). Use of soaking to enhance the bioavailability of iron and zinc from rice-based complementary foods used in the Philippines. *J Sci Food Agric.* 82:1115-1121.

122. Mustafa, K.D. and Adem, E. (2014). Comparison of autoclave, microwave, IR and UV-stabilization of whole wheat flour branny fractions upon the nutritional properties of whole wheat bread. *J Food Sci Tech.* 51(1):59-66.

123. Viteri, F.E., Alvarez, E. and Torun, B. (1983). Prevention of iron deficiency by means of iron fortification of sugar. In: Underwood B.A. ed. *Nutrition intervention strategies in national development*. Academic Press, New York. 287-314.

124. Viteri, F.E., Alvarez, E., Batres, R., Torún, B., Pineda, O., Mejía, L.A. and Sylvi, J. (1995). Fortification of sugar with iron sodium ethylenediamino-tetraacetate (FeNaEDTA) improves iron status in semirural Guatemalan populations. *Am J Clin Nutr.* 61:1153-63.

125. Ballot, D.E., MacPhail, A.P., Bothwell, T.H., Gillooly, M. and Mayet, F.G. (1989). Fortification of curry powder with NaFe(III)EDTA in an iron-deficient population: report of a controlled iron-fortification trial. *Am J Clin Nutr.* 49: 162-169.

126. Jorge, E.M., Wolfgang, H.P. and Peter, B. (2008). Biofortified crops to alleviate micronutrient malnutrition. *Curr Opin Plant Biol.* 11:166-170.

127. Pallauf, J., and Rimbach, G. (1997). Nutritional significance of phytic acid and *phytase*. *Arch Anim Nutr.* 50:301-319.

128. Cosgrove, D.J. (1966). The chemistry and biochemistry of inositol poly-phosphates. *Rev Pure Appl Chem*. 16:209-216.

129. Lasztity, R. and Lasztity, L. (1990). Phytic acid in cereal technology. In: Pomeranz Y. ed. *Advances in Cereal Science and Technology*. AACC, St. Paul, MN. pp. 309-371.

130. Shim, Y.H., Chae, B.J., and Lee, J.H. (2004). Effects of phytase and enzyme complex supplementation to diets with different nutrient levels on growth performance and ileal nutrient digestibility of weaned pigs. *Asian-Australas J Anim Sci*. 17(4):523-532.

131. Li, J., Hegeman, C.E., Hanlon, R.W., Lacy, G.H., Denbow, D.M. and Grabau, E.A. (1997). Secretion of active recombinant *phytase* from soybean cell-suspension cultures. *Plant Physiol*. 114:1103-1111.

132. Denbow, D.M., Ravindran, V., Kornegay, E.T., Yi, Z. and Hulet, R.M. (1995). Improving phosphorus availability in soybean meal for broilers by supplement -al *phytase*. *Poultry Sci.* 74:1831-1842.

133. Omogbenigun, F.O., Nyachoti, C.M. and Slominski, B.A. (2003). The effect of supplementing *phytase* and organic acids to a corn-soybean based diet fed to early-weaned pigs. *J Anim Sci.* 81:1806-1813.

134. Vats, P. and Banerjee, U.C. (2004). Production studies and catalytic properties of *phytases* (myo-inositol hexa*kis*phosphate phosphohydrolases): an overview. *Enzyme Microb Technol.* 35(1):3-14.

135. Sebastian, S., Touchburn, S.P., Chavez, E.R. and Lague, P.C. (1996). The effects of supplemental microbial phytase on the performance and utilization of dietary calcium, phosphorus, copper, and zinc in broiler chickens fed corn-soybean diets. *Poultry Sci.* 75:729-739.

136. Wodzinski, R.J., and Ullah, A.H.J. (1996). Phytase. *Adv Appl Microbiol.* 42:263-302.

137. Hughes, K.P. and Soares Jr. J.H. (1998). Efficacy of phytase on phosphorus utilization in practical diets fed to striped bass, *Morone saxatilis*. *Aquacult Nutr.* 4:133-140.

138. Phillippy, B.Q. (1999). Susceptibility of wheat and *Aspergillus niger phytases* to inactivation by gastrointestinal enzyme. *J Agric Food Chem*. 47:1385-1388.

139. Raboy V. (2007). The ABCs of low-phytate crops. *Nature Biotechnol*. 25(8): 874-875.

140. Wilcox, J.R., Premachandra, G.S., Young, K.A. and Raboy, V. (2000). Isolation of high seed inorganic P, low-phytate soybean mutants. *Crop Sci*. 40:1601-1605.

141. Oltmans, S.E., Fehr, W.R., Welke, G.A. and Cianzio, S.R. (2004). Inheritance of low-phytate phosphorus in soybean. *Crop Sci.* 44:433-435.

142. Yuan, F.J., H.J. Zhao, X.L. Ren, S.L. Zhu, X.J. Fu, and Q.Y. Shu. (2007). Generation and characterization of two novel low phytate mutations in soybean (*Glycine max* L. Merr.). *Theor Appl Genet.* 115:945-957.

143. Frank, T., Nörenberg, S. and Engel, K.H. (2009). Metabolite profiling of two novel low phytic acid (lpa) soybean mutants. *J Agric Food Chem.* 57:6408-6416.

144. Larson, S.R., Young, K.A., Cook, A., Blake, T.K. and Raboy, V. (1998). Linkage mapping of two mutations that reduce phytic acid content of barley grain. *Theor Appl Genet.* 97:141-141.

145. Ockenden, I., Dorsch, J.A., Reid, M.M., Lin, L., Grant, L.K., Raboy, V. and Loft, J.N.A. (2004). Characterization of the storage of phosphorus, inositol phosphate and cations in grain tissues of four barley (*Hordeum vulgare* L.) low phytic acid genotypes. *Plant Sci.* 167:1131-1142.

146. Dorsch, J.A., Cook, A., Young, K.A., Anderson, J.M., Bauman, A.T., Volkman, C.J., Murthy, P.N. and Raboy, V. (2003). Seed phosphorus and inositol phosphate phenotype of barley low phytic acid genotypes. *Phytochem.* 62:691-706.

147. Roslinsky, V., Eckstein, P.E., Raboy, V., Rossnagel, B.G. and Scoles, G.J. (2007). Molecular marker development and linkage analysis in three low phytic acid barley (*Hordeuni vulgare*) mutant lines. *Mol Breed.* 20:323-330.

148. Raboy, V., Gerbasi, P., Young, K., Stoneberg, S., Pickett, S., Bauman, A., Murthy, P., Sheridan, W. and Ertl, D. (2000). Origin and seed phenotype of maize *low phytic acid 1-1* and *low phytic acid 2-1*. *Plant Physiol.* 124:355-368.

149. Pilu, R., Panzeri, D., Gavazzi, G., Rasmussen, S.K., Consonni, C. and Nielsen, E. (2003). Phenotypic, genetic, and molecular characterization of a maize low phytic acid mutant (lpa241). *Theor Appl Genet.* 107:980-987.

150. Larson, S.R., Rutger, J.N., Young, K.A. and Raboy, V. (2000). Isolation and genetic mapping of a non-lethal rice (*Oryza sativa* L.) low phytic acid 1 mutation. Crop Sci. 40:1397-1405.

151. Liu, Q.L., Xu, X.H., Ren, X.L., Fu, H.W., Wu, D.X. and Shu, Q.Y. (2007). Generation and characterization of low phytic acid germplasm in rice (*Oryza sativa* L.). *Theor Appl Genet*. 114:803-814.

152. Kim, S.I., Andava, C.B., Goyal, S.S. and Tai, T.H. (2008). Isolation and characterization of a low phytic acid rice mutant reveals a mutation in the rice orthologue of maize MIK. *Theor Appl Genet*. 117.

153. Frank, T., Meuleye, B., Miller, A., Shu, Q. and Engel, K. (2007). Metabolite profiling of two low phytic acid (lpa) rice mutants. *J Agric Food Chem*. 55: 11011-11019.

154. Zhao, H., Liu, Q., Fu, H., Xu, X., Wu, D. and Shu, Q. (2008) Effect of non-lethal low phytic acid mutations on grain yield and seed viability in rice. *Field Crop Res.* 108:206-211.

155. Panzeri, D., Cassani, E., Doria, E., Tagliabue, G., Forti, L., Campion, B., Bollini, R., Brearley, C.A., Pilu, R., Nielsen, E. and Sparvoli F. (2011). A defective ABC transporter of the MRP family, responsible for the bean lpa1 mutation, affects the regulation of the phytic acid pathway, reduces seed myo-inositol and alters ABA sensitivity. *New Phytol.* 191:70-83.

156. Warkentin, T.D., Delgerjav T., Arganosa G., Rehman A.U., Bett K.E., Anbessa Y.K., Rossnagel B. and Raboy V. (2012). Development and characterization of low-phytate pea. *Crop Sci.* 52:74-78.

157. Meis, S., Fehr, W. and Schnebly, S. (2003). Seed source effect on field emergence of soybean lines with reduced phytate and raffinose saccharides. *Crop Sci.* 43:1336-1339.

158. Guttieri, M., Bowen, D., Dorsch, J., Raboy, V. and Souza, E. (2004). Identification and characterization of a low phytic acid wheat. *Crop Sci.* 44:418-424.

159. Pilu, R., Landoni, M., Cassani, E., Doria, E. and Nielsen, E. (2005). The maize lpa241 mutation causes a remarkable variability of expression and some pleiotropic effects. *Crop Sci.* 45:2096-2105.

160. Bregitzer, P. and Raboy, V. (2006). Effects of four independent low-phytate mutations in barley (*Hordeum vulgare* L.) on seed phosphorus characteristics and malting quality. *Cereal Chem.* 83:460-464.

161. Valluru, R. and van den Ende, W. (2011). Myo-inositol and beyond-Emerging networks under stress. *Plant Sci.* 181:387-400.

162. Gillaspy, G. (2011). The cellular language of myo-inositol signaling. New Phytol. 192:823-839.

163. Dieck, C., Boss, W. and Perera, I. (2012). A role for phosphoinositides in regulating plant nuclear functions. *Front Plant Sci.* 3.

164. Heilmann, M. and Heilmann, I. (2015). Plant phosphoinositides-complex networks controlling growth and adaption. *Biochim Biophys Acta.* 1851:759-769.

165. Pandey, A., Szakacs, G., Soccol, C.R., Rodriguez, Leon, J.A. and Soccol, A.T. (2001). Production, purification, and properties of microbial *phytases*. *Bio-resource Technol*. 77(3):203-214.

166. Pen, J., Verwoerd, T.C., van Paridon, P.A., Beudeker, R.J., van Elzen, P.J.M., Geerse, K., van der Klis, J.D., Versteegh, H.A.J., van Ooyen, A.J.J. and Hoekema, A. (1993). *Phytase* containing transgenic seeds as a novel feed additive for improved phosphorus utilization. *Biotechnol*. 11:811-814.

167. Drakakaki, G., Marcel, S., Glahn, R.P., Lund, E.K., Pariagh, S., Fisher, R., Christou, P., Stoger, E. (2005). Endosperm specific co-expression of recomb-inant soybean ferritin and *Aspergillus phytase* in maize results in significant increases in the levels of bioavailable iron. *Plant Mol Biol.* 59:869-880.

168. Chen, R., Xue, G., Chen, P., Yao, B., Yang, W., Ma, Q., Fan, Y., Zhao, Z., Tarczynski, M.C. and Shi J. (2008). Transgenic maize plants expressing a fungal *phytase* gene. *Transgenic Res.* 17(4):633-43.

169. Lucca, P., Hurrel, R. and Potrykus, I. (2001). Approaches to improving the bioavailability and level of iron in rice seeds. *Theor Appl Genet.* 102:392-397.

170. Brinch-Pedersen, H., Hatzack, F., Sorensen, L.D. and Holm, P.B. (2003). Concerted action of endogenous and heterologus *phytase* on phytic acid degradation in seed of transgenic wheat (*Triticum aestivum* L.). *Transgenic Res*. 12(6):649-659.

171. Hong, C.Y., Cheng, K.J., Tseng, T.H., Wang, C.S., Liu, L.F. and Yu, S.M. (2004). Production of two highly active bacterial *phytases* with broad pH optima in germinated transgenic rice seeds. *Transgenic Res.* 13:29-39.

172. Hamada, A., Yamaguchi, K., Harada, M., Horiguchi, K., Takahashi, T. and Honda, H. (2006). Recombinant, riceproduced yeast phytase shows the ability to hydrolyze phytate derived from seed-based feed, and extreme stability during ensilage treatment. *Biosci Biotechnol Biochem*. 70(6):1524-1527.

173. Chiera, J.M., Finer, J.J. and Grabau, E.A. (2004). Ectopic expression of a soybean *phytase* in developing seeds of *Glycine max* to improve phosphorus availability. *Plant Mol Biol.* 56:895-904.

174. Voinnet, O. (2002). RNA silencing: small RNAs as ubiquitous regulators of gene expression. *Curr Opin Plant Biol.* 5:444-451.

175. Matzke, M., Aufsatz, W., Kanno, T., Daxinger, L., Papp, I., Mette, M.F. and Matzke, A.J.M. (2004). Genetic analysis of RNA-mediated transcriptional gene silencing. *Biochim Biophys Acta*. 1677:129-141.

176. Brodersen, P. and Voinnet, O. (2006). The diversity of RNA silencing pathways in plants. *Trends Genet.* 22:268-280.

177. Vaucheret, H. (2006). Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes Dev*. 20:759-771.

178. Hannon, G.J. (2002). RNA interference. *Nature*. 418:244-251.

179. Grewal, S.I.S. and Moazed, D. (2003). Heterochromatin and epigenetic control of gene expression. *Science*. 301:798-802.

180. Baulcombe, D. (2004). RNA silencing in plants. *Nature*. 431:356-363.

181. Matzke, M., Kanno, T., Daxinger, L., Huettel, B. and Matzke, A.J.M. (2009). RNA-mediated chromatin-based silencing in plants. *Curr Opin Cell Biol.* 21:367-376.

182. Fabani, M.M. (2007). RNAi at Oxford. *J RNAi and Gene Silencing.* 3(1):220-222.

183. Antoni, R. and de Fougerolles. (2008). Delivery vehicles for small interfering RNA in vivo. *Hum Gene Ther.* 19:125-132.

184. Rahman, M., Ali, T.H. and Riazuddin, S. (2008). RNA interference: The story of gene silencing in plants and humans. *Biotechnol Adv.* 26(3):202-209.

185. Frizzi, A. and Huang S. (2010). Tapping RNA silencing pathways for plant biotechnology. *Plant Biotechnol J.* 8:655-677.

186. Kanazawa, A. (2008). RNA silencing manifested as visibly altered phenotypes in plants. *Plant Biotechnol.* 25:423-435.

187. Mansoor, S., Amin, I., Hussain, M., Zafar, Y. and Briddon, R.W. (2006). Engineering novel traits in plants through RNA interference. *Trends Plant Sci.* 11:559-565.

188. Napoli, C., Lemieux, C. and Jorgensen, R. (1990). Introduction of chimeric *chalcone synthase* gene into petunia results in reversible co-suppression of homologous genes in trans. *Plant Cell*. 2(4):279-89.

189. Romano, N. and Macino, G. (1992). Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. *Mol Microbiol*. 6:3343-3353.

190. Cogoni, C., Irelan, J.T., Schumacher, M., Schmidhauser, T., Selker, E.U. and Macino, G. (1996). Transgene silencing of the al-1 gene in vegetative cells of *Neurospora* is mediated by a cytoplasmic effector and does not depend on DNA-DNA interactions or DNA methylation. *EMBO J*. 15:3153-3163.

191. Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver S.E. and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*. 391:806-811.

192. Montgomery, M.K. and Fire, A. (1998). Double-stranded RNA as a mediator in sequence-specific genetic silencing and co-suppression. *Trends in Genetics*. 14:255-258.

193. Tabara, H., Grishok, A. and Mello, C.C. (1998). RNAi in C. elegans: Soaking in the genome sequence. *Science*. 282:430-431.

194. Guo, S. and Kemphues, K.J. (1995). par-1, a gene required for establishing poIarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell*. 81(4):611-20.

195. Waterhouse, P.M., Graham, M.W. and Wang, M.B. (1998). Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc Natl Acad Sci USA*. 95:13959-13964.

196. Hamilton, A.J. and Baulcombe, D.C. (1999). A species of small antisense RNA in post-transcriptional gene silencing in plants. *Science*. 286:950-952.

197. Stam, M., de Bruin, R., van Blokland, R., van der Hoorn, R.A., Mol, J.N. and Kooter, J.M. (2000). Distinct features of post-transcriptional gene silencing by antisense transgenes in single copy and inverted T-DNA repeat loci. *Plant J*. 21:27-42.

198. Van Houdt, H., Van Montagu, M. and Depicker, A. (2000). Both sense and antisense RNAs are targets for the sense transgene-induced post-transcript-ional silencing mechanism. *Mol Gen Genet*. 263:995-1002.

199. Tang, G. (2005). siRNA and miRNA: An insight into RISCs. *Trends Biochem Sci*. 30:106-114.

200. Bass, B.L. (2000). Double-stranded RNA as a template for gene silencing. Cell 101:235-238.

201. Fjose, A., Ellingsen, S., Wargelius, A. and Seo, H.C. (2001). RNA interference: mechanisms and applications. *Biotechnol Ann Rev.* 7:31-57.

202. Hammond, S.M., Caudy, A.A., and Hannon, G.J. (2001). Post-transcriptional gene silencing by double-stranded RNA. *Nat Rev Genet.* 2:110-119.

203. Willmann, M.R., Endres, M.W., Cook, R.T. and Gregory B.D. (2011). The Functions of RNA-Dependent RNA Polymerases in *Arabidopsis*. *Arabidopsis Book*, 9:e0146.

204. Bernstein, E., Caudy, A.A., Hammond, S.M. and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*. 409:363-366.

205. Szweykowska-Kulinska, Z., Jarmolowski, A., and Figlerowicz, M. (2003). RNA interference and its role in the regulation of eukaryotic gene expression. *Acta Biochim Polonica.* 50:217-229.

206. Du, Z., Lee, J.K., Tjhen, R., Stroud, R.M. and James, T.L. (2008). Structural and biochemical insights into the dicing mechanism of mouse Dicer: a conserved lysine is critical for dsRNA cleavage. *Proc Natl Acad Sci USA*. 105:2391-6.

207. Elbashir, S.M., Lendeckel, W. and Tuschl, T. (2001). RNA interference is mediated by 21 and 22 nt RNAs. *Genes Dev.* 15:188-200.

208. Tuschl, T. (2001). RNA interference and small interfering RNAs. *Chembio chem.* 2:239-45.

209. Hutvágner, G. and Zamore, P.D. (2002). RNAi: Nature abhors a double-strand. *Curr Opin Genetics Dev*. 12:225-232.

210. Sunkar, R. and Zhu, J.K. (2007). Micro RNAs and short-interfering RNAs in plants. *J Integr Plant Biol.* 49:817-826.

211. Watanabe, T. (2007). Analysis of small RNA profiles during development. *Meth Enzymol*. 427:155-169.

212. Li, J., Yang, Z., Yu, B., Liu, J. and Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3′-end uridylation activity in *Arabidopsis*. *Curr Biol.* 15:1501-1507.

213. Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L. and Hannon, G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. *Science*. 305:1437-1441.

214. Song, J.J., Smith, S.K., Hannon, G.J. and Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. *Science*. 305:1434-1437.

215. Schwarz, D., Tomari, Y. and Zamore, P. (2004). The RNA-induced silencing complex is a Mg-dependent *endonuclease*. *Curr Biol.* 14(9):787-791.

216. Ma, J., Yuan, Y., Meister, G., Pei, Y., Tuschl, T. and Patel, D. (2005). Structural basis for 5'-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature*. 434(7033):666-70.

217. Nykänen, A., Haley, B. and Zamore, P.D. (2001). ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell*. 107(3):309-21.

218. Parrish, S., Fleenor, J., Xu, S., Mello, C. and Fire, A. (2000). Functional anatomy of a dsRNA trigger; Differential requirement for the two trigger strands in RNA interference. *Mol Cell*. 6:1077-1087.

219. Zamore, P.D. (2006). RNA interference: Big applause for silencing in stockholm. *Cell*. 127:1083-1086.

220. Zamore, P.D. and Aronin, N. (2003). siRNAs knock down hepatitis. *Nat Med*. 9:266-267.

221. Lipardi, C., Wei, Q., Paterson and B.M. (2001). RNAi as random degradative PCR: siRNA primers convert mRNA into dsRNAs that are degraded to generate new siRNAs. *Cell*. 107(3):297-307.

222. Sijen, T., Fleenor, J., Simmer, F., Thijssen, K.L., Parrish, S., Timmons, L., Plasterk, R.H. and Fire, A. (2001). On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell*. 107(4):465-76.

223. Umesh, B.J., Ranjit, G.G. and Vishwas, A.B. (2011). Role of RNA interference in plant improvement. *Naturwissenschaften*. 98(6):473-492.

224. Tang, G. and Galili, G. (2004). Using RNAi to improve plant nutritional value: from mechanism to application. *Trends Biotechnol*. 22:463-469.

225. Liu, Q., Singh, S.P. and Green, A.G. (2002). High-stearic and high-oleic cottonseed oils produced by hairpin RNA-mediated post-transcriptional gene silencing. *Plant Physiol*. 129:1732-1743.

226. Eck, J.V., Conlin, B., Garvin, D.F., Mason, H., Navarre, D.A. and Brown, C.R. (2007). Enhanced beta-carotene content in potato via RNAi silencing of the beta-carotene *hydroxylase* gene. *Am J Potato Res*. 84:331-342.

227. Davuluri, G.R., van Tuinen, A., Fraser, P.D., Manfredonia, A., Newman, R., Burgess, D., Brummell, D.A., King, S.R., Palys, J., Uhlig, J., Bramley, P.M., Pennings, H.M. and Bowler, C. (2005). Fruit-specific RNAi-mediated suppression of *DET1* enhances carotenoid and flavonoid content in tomatoes. *Nature Biotechnol*. 23:890-895.

228. Yu, B., Lydiate, D.J., Young, L.W., Schäfer, U.A. and Hannoufa, A. (2008). Enhancing the carotenoid content of *Brassica napus* seeds by downregulating lycopene *epsilon cyclase*. *Transgenic Res*. 17(4):573-85.

229. Houmard, N.M., Mainville, J.L., Bonin, C.P., Huang, S., Luethy, M.H. and Malvar, T.M. (2007). High-lysine corn generated by endosperm-specific suppression of lysine catabolism using RNAi. *Plant Biotechnol J.* 5:605-614.

230. Regina, A., Bird, A., Topping, D., Bowden, S., Freeman, J., Barsby, T., Kosar-Hashemi, B., Li, Z., Rahman, S. and Morell, M. (2006). High-amylose wheat generated by RNA interference improves indices of large-bowel health in rats. *Proc Natl Acad Sci.* 103:3546-3551.

231. Shimada, T., Otani, M., Hamada, T. and Kim, S.H. (2006). Increase of amylose content of sweet potato starch by RNA interference of the starch branching enzyme II gene (*IbSBEII*). *Plant Biotechnol.* 23:85-90.

232. Regina, A., Kosar-Hashemi, B., Ling, S., Li, Z., Rahman, S. and Morell, M. (2010). Control of starch branching in barley defined through differential RNAi suppression of starch branching enzyme IIa and IIb. *J Exp Bot.* 61:1469-1482.

233. Ogita, S., Uefuji, H., Yamaguchi, Y., Koizumi, N. and Sano, H. (2003). RNA interference: producing decaffeinated coffee plants. *Nature*. 423:823.

234. Ogita, S., Uefuji, H., Morimoto, M. and Sano, H. (2004). Application of RNAi to confirm theobromine as the major intermediate for caffeine biosynthesis in coffee plants with potential for construction of decaffeinated varieties. *Plant Mol Biol*. 54:931-941.

235. Hüsken, A., Baumert, A., Milkowski, C., Becker, H.C., Strack, D. and Möllers, C. (2005). Resveratrol glucoside (Piceid) synthesis in seeds of transgenic oil seed rape (Brassica napus L.). *Theor Appl Genet*. 111:1553-1562.

236. Sunilkumar, G., Campbell, L.M., Puckhaber, L., Stipanovic, R.D. and Rathore, K.S. (2006). Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. *Proc Natl Acad Sci.* 103:18054-18059.

237. Kumar, P., Mundiyara, R. and Jitarwal, R.C. (2017). RNA Interference: New approach of gene silencing in plants. *Int. J. Pure App. Biosci*. 5(3): 421-425.

238. Jørgensen, K., Bak, S., Busk, P.K., Sørensen, C., Olsen, C.E., Puonti-Kaerlas, J. and Moller, B.L. (2005). Cassava plants with a depleted cyanogenic glucoside content in leaves and tubers. *Plant Physiol*. 139:363-374.

239. Shoemaker, R.C., Polzin, K., Labate, J., Specht, J., Brummer, E.C., Olson, T., Young, N., Concibido, V., Wilcox, J., Tamulonis, J.P., Kochert, G. and Boerma, H.R. (1996). Genome duplication in soybean (*Glycine* subgenus *soja*). *Genetics*. 144:329-338.

240. Shoemaker, R.C., Schlueter, J. and Doyle, J.J. (2006). Paleopolyploidy and gene duplication in soybean and other legumes. *Curr Opin Plant Biol.* 9:104-109.

241. Zeng, P., Vadnais, D., Zhang, Z. and Polacco, J. (2004). Refined glufosinate selection in *Agrobacterium*-mediated transformation of soybean [*Glycine max* (L.) Merr.]. *Plant Cell Rep.* 22:478-482.

242. Kasai, M. and Kanazawa, A. (2012). RNA silencing as a tool to uncover gene function and engineer novel traits in soybean. *Breed Sci*. 61(5):468.

243. Kinney, A.J. (1996). Development of genetically engineered soybean oils for food applications. *J Food Lipids*. 3:273-292.

244. Flores, T., Karpova, O., Su, X., Zeng, P., Bilyeu, K., Sleper, D.A., Nguyen, H.T. and Zhang, Z.J. (2008). Silencing of *GmFAD3* gene by siRNA leads to low α-linolenic acids (18:3) of fad3-mutant phenotype in soybean [*Glycine max* (Merr.)]. *Transgenic Res*. 17(5):839-850.

245. Wang, G. and Xu, Y. (2008). Hypocotyl-based *Agrobacterium*-mediated transformation of soybean (*Glycine max*) and application for RNA interference. *Plant Cell Rep*. 27(7):1177-1184.

246. Wagner, N., Mroczka, A., Roberts, P.D., Schreckengost, W. and Voelker, T. (2011). RNAi trigger fragment truncation attenuates soybean *FAD2-1* transcript suppression and yields intermediate oil phenotypes. *Plant Biotechnol J*. 9(7):723-728.

247. Yu, O., Shi, J., Hession, A.O., Maxwell, C.A., McGonigle, B. and Odell, J.T. (2003). Metabolic engineering to increase isoflavone biosynthesis in soybean seed. *Phytochemistry*. 63:753-763.

248. Lozovaya, V.V., Lygin, A.V., Zernova, O.V., Ulanov, A.V., Li, S., Hartman, G.L. and Widholm, J.M. (2007). Modification of phenolic metabolism in soybean hairy roots through down regulation of *chalcone synthase* or *isoflavone synthase*. *Planta*. 225:665-679.

249. Yi, J., Derynck, M.R., Li, X., Telmer, P., Marsolais, F. and Dhaubhadel, S. (2010). A single-repeat MYB transcription factor, GmMYB176, regulates *CHS8* gene expression and affects isoflavonoid biosynthesis in soybean. *Plant J*. 62:1019-1034.

250. Kinney, A.J., Jung, R. and Herman, E.M. (2001). Cosuppression of the α subunits of β-conglycinin in transgenic soybean seeds induces the formation of endoplasmic reticulum–derived protein bodies. *Plant Cell*. 13:1165-1178.

251. Schmidt, M.A., Barbazuk, W.B., Sandford, M., May, G., Song, Z., Zhou, W., Nikolau, B.J. and Herman, E.M. (2011). Silencing of soybean seed storage proteins results in a rebalanced protein composition preserving seed protein content without major collateral changes in the metabolome and transcriptome. *Plant Physiol*. 156:330-345.

252. Takagi, K., Nishizawa, K., Hirose, A., Kita, A. and Ishimoto, M. (2011). Manipulation of saponin biosynthesis by RNA interference-mediated silencing of *β-amyrin synthase* gene expression in soybean. *Plant Cell Rep*. 30(10):1835-1846.

253. Herman, E.M., Helm, R.M., Jung, R. and Kinney, A.J. (2003). Genetic modification removes an immunodominant allergen from soybean. *Plant Physiol*. 132(1):36-43.

254. Lee, M.Y., Shin, K.H., Kim, Y.K., Suh, J.Y., Gu, Y.Y., Kim, M.R., Hur, Y.S., Son, O., Kim, J.S., Song, E., Lee, M.S., Nam, K.H., Hwang, K.H., Sung, M.K., Kim, H.J., Chun, J.Y., Park, M., Ahn, T.I., Hong, C.B., Lee, S.H., Park, H.J., Park, J.S., Verma, D.P.S. and Cheon, C. (2005). Induction of thioredoxin is required for nodule development to reduce reactive oxygen species levels in soybean roots. *Plant Physiol*. 139:1881-1889.

255. Govindarajulu, M., Kim, S.Y., Libault, M., Berg, R.H., Tanaka, K., Stacey, G. and Taylor, C.G. (2009). *GS52* ecto-apyrase plays a critical role during soybean nodulation. *Plant Physiol*. 149:994-1004.

256. Libault, M., Zhang, X.C., Govindarajulu, M., Qiu, J., Ong, Y.T., Brechenmacher, L., Berg, R.H., Hurley-Sommer, A., Taylor, C.G. and Stacey, G. (2010). A member of the highly conserved *FWL* (tomato FW2.2-like) gene family is essential for soybean nodule organogenesis. *Plant J*. 62:852-864.

257. Chen, R., Matsui, K., Ogawa, M., Oe, M., Ochiai, M., Kawashima, H., Sakuradani, E., Shimizu, S., Ishimoto, M., Hayashi, M., Murooka, Y. and Tanaka, Y. (2006). Expression of Δ6, Δ5 desaturase and *GLELO elongase* genes from *Mortierella alpina* for production of arachidonic acid in soybean [*Glycine max* (L.) Merrill] seeds. *Plant Sci*. 170:399-406.

258. Schmidt, M.A. and Herman, E.M. (2008). Suppression of soybean oleosin produces micro-oil bodies that aggregate into oil body/ER complexes. *Mol Plant*. 1:910-924.

259. Dalton, D.A., Boniface, C., Turner, Z., Lindahl, A., Kim, H.J., Jelinek, L., Govindarajulu, M., Finger, R.E. and Taylor, C.G. (2009). Physiological roles of glutathione S-transferases in soybean root nodules. *Plant Physiol*. 150:521-530.

260. Arikit, S., Yoshihashi, T., Wanchana, S., Uyen, T.T., Huong, N.T., Wongpornchai, S. and Vanavichit, A., (2011). Deficiency in the amino aldehyde dehydrogenase encoded by *GmAMADH2*, the homologue of rice *Os2AP*, enhances 2-acetyl-1-pyrroline biosynthesis in soybeans (*Glycine max* L.). *Plant Biotechnol J*. 9:75-87.

261. Lee, J., Welti, R., Schapaugh, W.T. and Trick, H.N. (2011). Phospholipid and triacylglycerol profiles modified by PLD suppression in soybean seed. *Plant Biotechnol J*. 9:359-372.

262. Chuang, C. F. and Meyerowitz, E. M. (2000). Specific and heritable genetic interference by double-stranded RNA in *Arabidopsis thaliana*. *Proc Natl Acad Sci*. 97:4985-4990.

263. Wesley, S.V., Helliwell, C.A., Smith, N.A., Wang, M.B., Rouse, D.T., Liu, Q., Gooding, P.S., Singh, S.P., Abbott, D., Stoutjesdijk, P.A., Robinson, S.P., Gleave, A.P., Green, A.G. and Waterhouse, P.M. (2001). Construct design for efficient, effective and high throughput gene silencing in plants. *Plant J*. 27:581-590.

264. Kusaba, M. (2004). RNA interference in crop plants. *Curr Opin Biotechnol*. 15:139-143.

265. Watson, J.M., Fusaro, A.F., Wang, M. and Waterhouse, P.M. (2005). RNA silencing platforms in plants. *FEBS Lett*. 579:5982-5987.

266. Olhoft, P.M., Flagel, L.E., Donovan, C.M. and Somers, D.A. (2003). Efficient soybean transformation using hygromycin B selection in the cotyledonarynode method. *Planta*. 216:723-735.

267. Meurer, C.A., Dinkins, R.D. and Collins, G.B. (1998). Factors affecting soybean cotyledonary node transformation. *Plant Cell Rep*. 18:180-186.

268. DeBlock, M., Botterman, J., van de Wiele, M., Dockx, J., Thoen, C., Gossele, V., Rao Movva, N., Thompson, C., van Montagu, M. and Leemans, J. (1987). Engineering herbicide resistance in plants by expression of a detoxifying enzymes. *EMBO J*. 6:2513-2518.

269. Christou, P., Ford, T.L. and Kofron, M. (1991). Production of transgenic rice (Oryza sativa L.) plants from agronomically important indica and japonica varieties via electric discharge particle acceleration of exogenous DNA into immature zygotic embryos. *Biotechnol*. 9:957-962.

270. Vasil, V., Castillo, A.M., Fromm, M.E. and Vasil, I.K. (1992). Herbicide resistant fertile transgenic wheat plants obtained by microprojectile bombardment of regenerable embryogenic callus. *Nature Biotechnol*. 10(6):667-674.

271. Zhao, Z.Y., Cai, T., Tagliani, L., Miller, M., Wang, N., Pang, H., Ridert, R., Schroder, S., Hondred, D., Seltzer, J. and Pierce, D. (2000). *Agrobacterium*-mediated sorghum transformation. *Plant Mol Biol*. 44:789-798.

272. Emani, C., Sunilkumar, G. and Rathore, K.S. (2002). Transgene silencing and reactivation in sorghum. *Plant Sci*. 162:181-192.

273. Raboy, V., and Gerbasi, P. (1996). Genetics of myo-inositol synthesis and accumulation. In: Biswas B.B., Biswas S. eds. *Subcellular Biochemistry, Vol. 26: myo-Inositol Phosphates, Phosphoinositides, and Signal transduction*. Plenum Press, New York. pp. 257-285.

274. Shi, J., Wang, H., Hazebroek, J., Ertl, D.S. and Harp, T. (2005). The maize low-phytic acid 3 encodes a myo-inositol kinase that plays a role in phytic acid biosynthesis in developing seeds. *Plant J*. 42:708-719.

275. Asada, K., Tanaka, K. and Kasai, Z. (1969). Formation of phytic acid in cereal grains. *Ann N Y Acad Sci*. 165:801-814.

276. Drøbak, B.K. (1992). The plant phosphoinositide system. *Biochem J*. 288, 697 - 712.

277. Frederick, J.P., Mattiske, D., Wofford, J.A., Megosh, L., Drake, L.Y., Chiou, S.T., Hogan, B.L. and York, J.D. (2005). An essential role for a nuclear *inositol polyphosphate multikinase*, Ipk2, in mouse embryogenesis and second messenger production. *Proc Natl Acad Sci*. 102(24):8454-9.

278. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular cloning: a laboratory manual, 2nd ed. CSHL Press, Cold Spring Harbor, N.Y.

279. Sugimoto, T., Kawasaki, T., Kato, T., Whittier, R.F., Shibata, D. and Kawamura, Y. (1992). cDNA sequence and expression of a *phosphoenolpyruvate carboxylase* gene from soybean. *Plant Mol Biol*. 20:743-747.

280. Tuteja, J.H., Clough, S.J., Chan, W.C. and Vodkin, L.O. (2004). Tissue-specific gene silencing mediated by a naturally occurring *chalcone synthase* gene cluster in *Glycine max*. *Plant Cell*. 16(4):819-835.

281. Livak, K.J. and Schmittgen, T.D. (2001). Analysis of relative gene expression data using realtime quantitative PCR and the 2(-ΔΔC(T)) method. *Methods*. 25(4):402-408.

282. Rice, P., Longden, I., Bleasby, A. (2000). Emboss: The European molecular biology open software suite. *Trends Genet*. 16:276-277.

283. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook*. pp. 571-607.

284. Emanuelsson, O., Nielsen, H., Brunak, S. and Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Mol Biol*. 300:1005-1016.

285. Chou, K.C. and Shen, H.B. (2007) MemType-2L: A Web server for predicting membrane proteins and their types by incorporating evolution information through Pse-PSSM. *Biochem Biophys Res Commun*. 360:339-345.

286. Hua, S. and Sun, Z. (2001). Support vector machine approach for protein subcellular localization prediction. *Bioinformatics*. 17:721-728.

287. Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J. and Nakai, K. (2007). WoLF PSORT: protein localization predictor. *Nucl Acids Res*. 35:W585-587.

288. Yu, C.S., Chen, Y.C., Lu, C.H. and Hwang, J.K. (2006). Prediction of protein subcellular localization. *Proteins*. 64:643-651.

289. Buchan, D.W.A., Minneci, F., Nugent, T.C.O., Bryson, K. and Jones, D.T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res*. 41:W340-W348.

290. Hofmann, K. and Stoffel, W. (1993). TMbase - A database of membrane spanning proteins segments. *Biol Chem*. 374:166.

291. Rost, B., Casadio, R. and Fariselli, P. (1996). Refining neural network predictions for helical transmembrane proteins by dynamic programming. *Proc Int Conf Intell Syst Mol Biol*. 4:192-200.

292. Petersen, T.N., Brunak, S., Heijne, G. and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*. 8:785-786.

293. Wallace, I.M., Sullivan, O., Higgins, D.G. and Notredame, C. (2006). M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res*. 34:1692-1699.

294. Zuckerkandl, E. and Pauling, L. (1965). Evolutionary divergence and convergence in proteins. In: Bryson V., Vogel H.J. eds. *Evolving Genes and Proteins*. Academic Press, New York. pp. 97-166.

295. Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 30:2725-2729.

296. Letunic, I. and Bork, P. (2016). Interactive Tree of Life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucl Acids Res*. 44:W242-W245.

297. Lescot, M., Dehais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouze, P. and Rombauts, S. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res*. 30:325-327.

298. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. and Barton, G.J. (2009). Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 25(9):1189-1191.

299. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W., Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*. 37:W202-W208.

300. Pagni, M., Ioannidis, V., Cerutti, L., Zahn-Zabal, M., Jongeneel, C.V., Hau, J., Martin, O., Kuznetsov, D. and Falquet, L. (2007). MyHits: improvements to an interactive resource for analyzing protein sequences. *Nucleic Acids Res*. 35: W433-W437.

301. Geourjon, C. and Deléage, G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci*. 11:681-684.

302. Rost, B. and Sander, C. (1993). Prediction of protein secondary structure at better than 70% accuracy. *Mol Biol*. 232:584-599.

303. Ferrè, F. and Clote, P. (2005). DiANNA: a web server for disulfide connectivity prediction. *Nucleic Acids Res*. 33:W230-2.

304. Hutchinson, E.G. and Thornton, J.M. (1996). PROMOTIF--a program to identify and analyze structural motifs in proteins. *Protein Sci*. 5(2):212-220.

305. Bond, C.S. (2003), TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics*. 19:311-2.

306. Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L. and Schwede, T. (2014). SWISS-MODEL: modeling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*. 42:W252-W258.

307. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protocol*. 10:845-858.

308. Webb. B and Sali, A. (2014). Comparative Protein Structure Modeling Using Modeller. *Curr Protoc Bioinformatics* 5.6.1-5.6.32.

309. Gabrielsen, M., Ravna, A.W., Kristiansen, K. and Sylte, I., (2011), Substrate binding and translocation of the serotonin transporter studies by docking and molecular dynamics simulations, *J Mol Model*. Doi: 10.1007/s00894-011-1133-1.

310. Melo, F., Sanchez, R. and Sali, A. (2002). Statistical Potentials for Fold Assessment. *Protein Sci*. 11(2): 430-448.

311. Shen, M. and Sali, A. (2006). Statistical Potential for Assessment and Prediction of Protein Structures. *Protein Sci*. 15(11): 2507-2524.

312. Lovell, S.C., Davis, I.W., Arendall III, W.B., de Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S. and Richardson, D.C. (2003). Structure validation by Cα geometry: Φ, Ψ, and Cβ deviation. *Proteins*. 50:437-450.

313. Eisenberg, D., Lüthy, R. and Bowie, J.U. (1997). VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol*. 277:396-404.

314. Wiederstein, M. and Sippl, M.J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 35:W407-W410.

315. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004). UCSF Chimera: a Visualization System for Exploratory Research and Analysis. *J Comput Chem*. 25(13):1605-1612.

316. Kawabata, T. (2003). MATRAS: a program for protein 3D structure comparison. *Nucleic Acids Res*. 31:3367-9.

317. Ishida, T. and Kinoshita, K. (2007). PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res*. 35.

318. Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M.R., Smith, J.C., Kasson, P.M., Spoel, D., Hess, B. and Lindahl, E. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*. 29:845-854.

319. Turner, P.J. (2005). XMGRACE, Version 5.1.19. Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton, OR.

320. Abagyan, R., Totrov, M., Kuznetsov, D. (2004). ICM - A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J Comput Chem*. 15 (5): 488-506.

321. Bursulaya, B.D., Totrov, M., Abagyan, R. and Brooks III, C.L. (2003). Comparative study of several algorithms for flexible ligand docking. *J Comput Aided Mol Des*. 17:755-763.

322. Leach, A.R., Schichet, B.K. and Peishoff, C.E. (2006). Prediction of Protein-Ligand Interactions. Docking and Scoring: Successes and Gabs, *J Med Chem.* 49(20).

323. Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y. and Liang, J. (2006). CASTp: computed atas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res*. 34:W116-W118.

324. Kozakov, D., Grove, L.E., Hall, D.R., Bohnuud, T., Mottarella, S.E., Luo, L., Xia, B., Beglov, D. and Vajda, S. (2015). The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat Protocols*. 10:733-755.

325. Roche, D.B., Buenavista, M.T. and McGuffin, L.J. (2013). The FunFOLD2 server for the prediction of protein-ligand interactions. *Nucleic Acids Res*. 41:W303-307.

326. Höltje, H.D., Sippl, W., Rognan, D. and Folkers, G. (2003). Molecular modeling. *Basic principles and applications*, 2nd Edition. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany. pp. 15-25, 59-73, 87-110, 146-165.

327. Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N. and Rokhsar, D.S. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*. 40:D1178-D1186.

328. Franklin-Tong, V.E., Drobak, B.K., Allan, A.C. and Trewavas, A.J. (1996). Growth of pollen tubes of *Papaver rhoeas* is regulated by a slow-moving calcium wave propagated by inositol 1,4,5-trisphosphate. *Plant Cell*. 8:1305-1321.

329. Malho, R. (1998). Role of 1,4,5-inositol trisphosphate-induced Ca2+ release in pollen tube orientation. *Sex Plant Reprod*. 11:231-235.

330. Pierson, E.S., Miller, D.D., Callaham, D.A., Shipley, M.A., Rivers, B.A., Cresti, M. and Hepler, P.K. (1994). Pollen tube growth is coupled to the extracellular calcium ion flux and the intracellular calcium gradient: effect of BAPTA-type buffers and hypertonic media. *Plant Cell*. 6:1815-1828.

331. Xu, J., Brearley, C.A., Lin, W.H., Wang, Y., Ye, R., Mueller-Roeber, B., Xu, Z.H. and Xue, H.W. (2005). A Role of *Arabidopsis* inositol polyphosphate kinase, *AtIPK2α*, in Pollen Germination and Root Growth. *Plant Physiol*. 137:94-103.

332. Bibikova, T.N., Zhigiler, A. and Gilroy, S. (1997). Root hair growth in *Arabidopsis thaliana* is directed by calcium and an endogenous polarity. *Planta*. 203:495-505.

333. Felle, H.H. and Hepler, P.K. (1997). The cytosolic Ca2+ concentration gradient of Sinapis alba root hairs as revealed by Ca2+ selective microelectrode tests and fura-dextran ratio imaging. *Plant Physiol*. 114:39-45.

334. Wymer, C.L., Bibikova, T.N. and Gilroy, S. (1997). Cytoplasmic free calcium distribution during the development of root hairs of *Arabidopsis thaliana*. *Plant J*. 12:427-439.

335. Zhang, Z.B., Yang, G., Arana, F., Chen, Z., Li, Y. and Xia, H.J. (2007). *Arabidopsis inositol polyphosphate 6-/3-kinase* (AtIpk2β) is involved in axillary shoot branching via auxin signaling. *Plant Physiol*. 144:942-951.

336. Raboy, V. and Dickinson, D.B. (1987). The timing and rate of phytic acid accumulation in developing soybean seeds. *Plant Physiol*. 85:841-844.

337. Takaiwa, F., Yamanouchi, U., Yoshihara, T., Washida, H., Tanabe, F., Kato, A. and Yamada, K. (1996). Characterization of common cis-regulatory elements responsible for the endosperm-specific expression of members of the rice glutelin multigene family. *Plant Mol Biol*. 30:1207-1221.

338. Onodera, Y., Suzuki, A., Wu, C.Y., Washida, H. and Takaiwa, F. (2001). A rice functional transcriptional activator, RISBZ1, responsible for endosperm-specific expression of storage protein genes through GCN4 motif. *J Biol Chem*. 276:14139-14152.

339. Washida, H., Wu, C.Y., Suzuki, A., Yamanouchi, U., Akihama, T., Harada, K. and Takaiwa, F. (1999). Identification of cis-regulatory elements required for endosperm expression of the rice storage protein glutelin gene *GluB-1*. *Plant Mol Biol*. 40:1-12.

340. Yamaguchi-Shinozaki, K., Mundy, J. and Chua, N.H. (1989). Four tightly linked rab genes are differentially expressed in rice. *Plant Mol Biol*. 14:29-39.

341. Mongkolsiriwatana, C., Pongtongkam, P. and Peyachoknagul, S. (2009). In silico promoter analysis of photoperiod-responsive genes identified by DNA microarray in rice (Oryza sativa L.). *Kasetsart J (Nat Sci)*. 43:164-177.

342. Niu, J., Wang, J., Hu, H., Chen, Y., An, J., Cai, J., Sun, R., Sheng, A., Liu, X. and Lin, S. (2016). Cross-talk between freezing response and signalling for regulatory transcriptions of *MIR475b* and its targets by miR475b promoter in Populous suaveolens. *Sci Rep*. 6:20648.

343. Lois, R., Dietrich, A., Hahlbrock, K. and Schulz, W. (1989). A phenylalanine ammonia-lyase gene from parsley: structure, regulation and identification of elicitor and light responsive cis-acting elements. *EMBO J*. 8:1641-1648.

344. Arguello-Astorga, G.R. and Herrera-Estrella, L.R. (1996). Ancestral multipartite units in light-responsive plant promoters have structural features correlating with specific phototransduction pathways. *Plant Physiol*. 112:1151-1166.

345. Sessa, G., Meller, Y. and Fluhr, R. (1995). A GCC element and a G-box motif participate in ethylene-induced expression of the *PRB-1b* gene. *Plant Mol Biol*. 28:145-153.

346. Feldbrugge, M., Hahlbrock, K. and Weisshaar, B. (1996). The transcriptional regulator CPRF1: expression analysis and gene structure. *Mol Gen Genet*. 251:619-627.

347. Kim, S.R., Kim, Y. and An, G. (1993). Identification of methyl jasmonate and salicylic acid response elements from the nopaline synthase (nos) promoter. *Plant Physiol*. 103:97-103.

348. Rouster, J., Leah, R., Mundy, J. and Cameron-Mills, V. (1997). Identification of a methyl jasmonate-responsive region in the promoter of a *lipoxygenase 1* gene expressed in barley grain. *Plant J*. 11:513-523.

349. Shinozaki, K. and Yamaguchi-Shinozaki, K. (2000). Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. *Curr Opin Plant Biol*. 3:217-223.

350. Abe, H., Urao, T., Ito, T., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2003). *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signalling. *Plant Cell*. 15:63-78.

351. Nguyen, T., Sherratt, P.J., Cecil, B. and Pickett, C.B. (2003). Regulatory mechanisms controlling gene expression mediated by the antioxidant response element. *Annu Rev Pharmacol Toxicol*. 43:233-260.

352. Matsuno, K. and Fujimura, T. (2014). Induction of phytic acid synthesis by abscisic acid in suspension-cultured cells of rice. *Plant Sci*. 217-218:152-157.

353. Aggarwal, S., Shukla, V., Bhati, K.K., Kaur, M., Sharma, S., Singh, A., Mantri, S. and Pandey, A.K. (2015). Hormonal Regulation and Expression Profiles of Wheat Genes Involved during Phytic Acid Biosynthesis Pathway. Plants (Basel), 4:298-319.

354. Gill, S.C. and Von Hippel, P.H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem*. 182:319-326.

355. Guruprasad, K., Reddy, B.V.B. and Pandit, M.W. (1990). Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng*. 4:155-161.

356. Rogers, S., Wells, R. and Rechsteiner, M. (1986). Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science*. 234:364-368.

357. Ikai, A.J. (1980). Thermostability and aliphatic index of globular proteins. *J Biochem*. 88:1895-1898.

358. Kyte, J. and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol*. 157:105-132.

359. Schwartz, R., Ting, C.S. and King, J. (2001). Whole proteome pI values correlate with subcellular localizations of proteins for organisms within the three domains of life. *Genome Res*. 11(5):703-9.

360. Pashou, E.E., Litou, Z.I., Liakopoulos, T.D. and Hamodrakas, S.J. (2004). WaveTM: wavelet-based transmembrane segment prediction. *In Silico Biol*. 4:127-31.

361. Saiardi, A., Erdjument-Bromage, H., Snowman, A., Tempst, P. and Snyder, S.H. (1999). Synthesis of diphosphoinositol pentakisphosphate by a newly identified family of higher inositol polyphosphate kinases. *Curr Biol*. 9:1323-1326.

362. Holmes, W. and Jogl, G. (2006). Crystal structure of inositol phosphate multikinase 2 and implications for substrate specificity. *J. Biol Chem*. 281:38109-38116.

363. Cooper, J.A., Esch, F.S., Taylor, S.S. and Hunter, T. (1984). Phosphorylation sites in enolase and lactate dehydrogenase utilized by tyrosine protein kinases in vivo and in vitro. *Biol Chem*. 259:7835-7841.

364. Steinberg, S.F. (2008). Structural basis of protein kinase C isoform function. *Physiol Rev*. 88:1341-1378.

365. Nishikawa, K., Toker, A., Johannes, F.J., Songyang, Z. and Cantley, L.C. (1997). Determination of the specific substrate sequence motifs of protein kinase C isozymes. *Biol Chem*. 272:952-960.

366. Neumann, G.M., Thomas, I. and Polya, G.M. (1996). Identification of the site on potato carboxypeptidase inhibitor that is phosphorylated by plant calcium-dependent protein kinase. *Plant Sci*. 114:45-51.

367. Roberts, D.M. and Harmon, A.C. (1992). Calcium modulated proteins: targets of intracellular calcium signals in higher plants. *Annu Rev Plant Physiol and Plant Mol Biol*. 43:375-414.

368. Josefsen, L., Bohn, L., Sorensen, M. and Rasmussen, S. (2007). Characterization of a multifunctional inositol phosphate kinase from rice and barley belonging to the ATP-grasp superfamily. *Gene*. 397:114-125.

369. Plevin, M.J., Mills, M.M. and Ikura, M. (2005). The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. *Trends Biochem Sci*. 30:66-69.

370. Agarwal, R., Mumtaz H., Ali N. (2009) Role of inositol polyphosphates in programmed cell death. *Mol Cell Biochem*. 328:155-165.

371. Irvine, R.F. and Schell, M.J. (2001). Back in the water: the return of the inositol phosphates. *Nat Rev Mol Cell Biol*. 2:327-338.

372. Shears, S.B. (2004). How versatile are inositol phosphate kinases? *Biochem J*. 377:265-280.

373. Reehana, N., Ahamed, A.P., Ali, D.M., Suresh, A., Kumar, R.A., Thajuddin, N. (2013). Structure based computational analysis and molecular phylogeny of C-Phycocyanin gene from the selected cynobacteria. *International J Biological, Veterinary, Agric. Food Eng*. 7:47-51.

374. Alber, T., Gilbert, W.A., Ponzi, D.R. and Petsko, G.A. (1983). The role of mobility in the substrate binding and catalytic machinery of enzymes. *Ciba Found Symp*. 93:4-24.

375. Dunker, A.K., Brown, C.J., Lawson, J.D., Iakoucheva, L.M. and Obradović, Z. (2002). Intrinsic disorder and protein function. *Biochem*. 41:6573-6582.

376. Klaholz, B. and Moras, D. (2002). C-H...O hydrogen bonds in the nuclear receptor RAR gamma-a potential tool for drug selectivity. *Structure*. 10:1197-1204.

377. Lott J.N.A., Greenwood, J.S., Batten, G.D. (1995). Mechanisms and regulation of mineral nutrient storage during seed development. In: Kigel J., Galili G. eds. *Seed Development and Germination*. Marcel Dekker Inc., New York. pp. 215-235.

378. Mendoza, C. (2002). Effect of genetically modified low phytic acid plants on mineral absorption. *Int J Food Sci Technol*. 37:759-767.

379. Schlemmer, U., Frølich, W., Prieto, R.M., Grases, F. (2009). Phytate in foods and significance for humans: Food sources, intake, processing, bioavailability, protective role and analysis. *Mol Nutr Food Res*. 53:S330-S375.

380. Leytem, A.B., Maguire, R.O. (2007). Environmental implications of inositol phosphates in animal manures. In: Turner B.L., Richardson A.E., Mullaney E.J. eds. *Inositol Phosphates: Linking Agriculture and the Environment*. CAB International, Wallingford, UK. pp. 150-168.

381. Rasmussen, S., and F. Hatzack. (1998). Identification of two low-phytate barley (Hordeum vulgare L.) grain mutants by TLC and genetic analysis. *Hereditas*. 129:107-112.

382. Mendoza, C., Viteri, V.E., Lo¨nnerdal, B., Young, K.A., Raboy, V., Brown, K.H. (1998). Effect of genetically modified, low-phytic acid maize on absorption of iron from tortillas. *Am J Clin Nutr*. 68:1123-1128.

383. K.M. Hambidge, J.W. Huffer, V. Raboy, G.K. Grunwald, J.L. Westcott, L. Sian, L.V. Miller, J.A. Dorsch, N.F. Krebs, Zinc absorption from low-phytate hybrids of maize and their wild-type hybrids. *Am. J. Clin. Nutr*. 79(2004) 1053-1059.

384. Hambidge, K.M., Krebs, N.F., Westcott, J.L., Sian, L., Miller, L.V., Peterson, K.L., Raboy, V. (2005). Absorption of calcium from tortilla meals prepared from low-phytate maize. *Am J Clin Nutr*. 82:84-87.

385. Cichy, K.A. and Raboy, V. (2008). Evaluation and development of low phytate crops. In: Krishnan H. ed. *Modification of Seed Composition to Promote Health and Nutrition, Agronomy Monograph*, Vol. 51. American Society of Agronomy, and Crop Science Society of America. pp. 177-200.

386. Hegeman, C.E., and Grabau, E.A. (2001). A novel phytase with sequence similarity to purple acid phosphatases is expressed in cotyledons of germinating soybean seedlings. *Plant Physiol*. 126:1598-1608.

387. Ali, N., Paul, S., Gayen, D., Sarkar, S.N., Datta, S.K., and Datta, K. (2013). RNAi mediated down regulation of myo-inositol-3-phosphate synthase to generate low phytate rice. *Rice*. 6:12.

388. Ali, N., Paul, S., Gayen, D., Sarkar, S.N., Datta, K., and Datta, S.K. (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: e68161.

389. Bhati, K.K., Alok, A., Kumar, A., Kaur, J., Tiwari, S., and Pandey, A.K. (2016). Silencing of ABCC13 transporter in wheat reveals its involvement in grain development, phytic acid accumulation and lateral root formation. *J Exp Bot*. 67(14), 4379-4389 (2016).

390. Wen, C.M., Cai, P.Z., Xu, Z.J. and Zhong, W.F. (2008). A Novel Approach Obtaining Intron-Containing Hairpin RNA Constructs. *Biosci Biotechnol Bio chem*. 72(2): 615-617.

391. Manjul, D., Sadanand, A.D., Leonardo, S., Raju, K., Jude, W.G. (2014). Temporal and spatial control of gene expression in horticultural crops. *Hortic Res*. 1:14047.

392. Kasuga, M., Liu, Q., Miura, S., Yamaguchi-Shinozaki, K., Shinozaki, K. (1999). Improving plant drought, salt, and freezing tolerance by gene transfer of a single stress-inducible transcription factor. *Nat Biotechnol*. 17:287-291.

393. Hsieh, T.H., Lee, J.T., Yang, P.T. et al. (2002). Heterology expression of the *Arabidopsis* C-repeat/dehydration response element binding factor 1 gene confers elevated tolerance to chilling and oxidative stresses in transgenic tomato. *Plant Physiol*. 129:1086-1094.

394. Karnwal, M.K. and Singh K. (2009). Studies on genetic variability, character association and path coefficient for seed yield and its contributing traits in soybean [*Glycine max* (L.) Merrill]. *Legume Research*. 32:70-73.

395. Ramteke, R., Kumar, V., Murlidharan, P. and Agarwal, D.K. (2010). Study on genetic variability and traits interrelationship among released soybean varieties of India [*Glycine max* (L.) Merrill]. *Electronic J of Plant Breeding*. 1,

1483-1487.

396. Ramteke, R., and Murlidharan, P. (2012). Characterization of soybean (*Glycine max*) varieties as per DUS guidelines. *Indian J of Agric Sci*. 82:572-577.

397. Hood, E.E., Gelvin, S.B., Melchers, L.S. and Hoekema, A. (1993). New *Agrobacterium* helper plasmids for gene transfer to plants. *Transgenic Research*. 2:208-218.

398. Paz, M.M., Shou, H., Guo, Z., Zhang, Z., Banerjee, A.K., Wang, K. (2004). Assessment of conditions affecting *Agrobacterium*-mediated soybean transformation using the cotyledonary node explant. *Euphytica*. 136:167-179.

399. Gamborg, O.L., Miller, R.A. and Ojima, K. (1968). Nutrient requirements of suspension cultures and soybean root cells. *Exp Cell Res*. 50: 151-158.

400. Hinchee, M.A.W., Ward, D.V.C., Newell, C.A., McDonnell, R.E., Sato, S.J., Gasser, C.S., Fischhoff, D.A., Re, D.B., Fraley, R.T. and Horsch, R.B. (1988). Production of transgenic soybean plants using *Agrobacterium*-mediated DNA transfer. *Bio Technol*. 6:915-922.

401. Di, R., Purcell, V., Collins, G.B., Ghabrial, S.A. (1996). Production of transgenic soybean lines expressing the bean pod mottle virus coat protein precursor gene. *Plant Cell Rep*. 15:746-750.

402. Zhang, Z., Coyne, D.P. and Mita, A. (1997). Factors affecting *Agrobacterium*-mediated transformation of common bean. *J Amer Soc Hort Sci*. 122:300-305.

403. Murashige, T. and Skoog, F. (1962). A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol Plant*. 15:473-479.

404. Murray, M.G. and Thompson, W.F. (1980). Rapid Isolation of High Molecular Weight Plant DNA. *Nucleic Acids Research*. 8:4321-4326.

405. Lehrfeld J. (1989). High-performance liquid chromatography analysis of phytic acid on a pH-stable, macroporous polymer column. *Cereal Chem*. 66: 510-515.

406. Pandey, V., Krishnan, V., Basak, N., Hada, A., Punjabi, M., Jolly, M., Lal, S.K., Singh, S.B., Sachdev, A. (2016). Phytic acid dynamics during seed development and it's composition in yellow and black Indian soybean (*Glycine max* L.) genotypes through a modified extraction and HPLC method. *J Plant Biochem and Biotechnol*. 25, 367.

407. Dost, K. and Tokul, O. (2006). Determination of phytic acid in wheat and wheat products by reverse phase high performance liquid chromatography. *Anal Chim Acta*. 558:22-27.

408. Chen, P.S., Toribara, T.Y., and Warner, H. (1956). Microdetermination of Phosphorus. *Analytical Chemistry*. 28:1756-1758.

409. Al-Amery, M., Fukushige H., and Hildebrand D. (2015). Single seed selection for low phytate lines. *J Am Oil Chem Soc*. 92:1119-1123.

410. Kiers, J.L., Nout, M.J.R., and Rombouts, F.M. (2000). In vitro digestibility of processed and fermented soya bean, cowpea and maize. *J Sci Food Agric.* 80:1325-1331.

411. Campion, B., Sparvoli, F., Doria, E., Tagliabue, G., Galasso, I., Fileppi, M., Bollini, R., and Nielsen, E. (2009). Isolation and characterisation of a lpa (low phytic acid) mutant in common bean (*Phaseolus vulgaris* L.). *Theor Appl Genet*. 118:1211-21.

412. Kaneda, Y., Tabei, Y., Nishimura, S., Harada, K., Akihama, T. and Kitamura, K. (1997). Combination of thidiazuron and basal media with low salt concentrations increases the frequency of shoot organogenesis in soybeans (*Glycine max* L. Merr.). *Plant Cell Rep*. 17:8-12.

413. Gahakwa, D., Maqbool, S.B., Fu, X., Sudhakar, D., Christou, P., Kohli, A. (2000). Transgenic rice as a system to study the stability of transgene expression: multiple heterologous transgenes show similar behavior in diverse genetic backgrounds. *Theor Appl Genet*. 101:388-399.

414. Hadi, F., Salmanian, A.H., Mousavi, A., Ghazizadeh, E., Amani, J., Noghabi, K.A. (2012). Development of quantitative competitive PCR for determination of copy number and expression level of the synthetic glyphosate oxido-reductase gene in transgenic canola plants. *Electron J Biotechn*. 15(4):2-14.

415. Tubana, B.S. and George, T. (1997). Seed phosphorus fertilizer effect on the early growth of soybean under low phosphorus supply. *Philipp J Crop Sci.* 22:32.

416. Horsley, S.B. and Gottschalk, K.W. (1993). Leaf area and net photosynthesis during development of *Prunus serotina* seedlings. *Tree Physiol*. 12:55-69.

417. Adams, C.L., Hambidge, M., Raboy, V., Dorsch, J.A., Sian, L., Westcott, J.L., and Krebs N.F. (2002). Zinc absorption from a low-phytic acid maize. *Am J Clin Nutr*. 76:556-559.

418. Linares, L.B., Broomhead, J.N., Guaiume, E.A., Ledoux, D.R., Veum, T.L., and Raboy, V. (2006). Effects of low phytate barley (Hordeum vulgare L.) on zinc utilization in young broiler chicks. *Poult Sci*. 86:299-308.

419. Mazariegos, M., Hambidge, K.M., Krebs, N.F., Westcott, J.E., Lei, S., Grunwald, G.K., Campos, R., Barahona, B., Raboy, V., and Solomons, N.W. (2006). Zinc absorption in Guatemalan school children fed normal or low-phytate maize. *Am J Clin Nutr*. 83:59-64.

420. Veum, T.L., Ledoux, D.R., and Raboy, V. (2007). Low-phytate barley cultivars improve the utilization of phosphorus, calcium, nitrogen, energy, and dry matter in diets fed to young swine. *J Anim Sci*. 85:961-971.

421. Persson, D.P., Hansen, T.H., Laursen, K.H., Schjoerring, J.K., and Husted, S. (2009). Simultaneous iron, zinc, sulfur and phosphorus speciation analysis of barley grain tissues using SEC-ICP-MS and IP-ICP-MS. *Metallomics*. 1:418-426.

Appendix I

Figures

Figure 1: Prediction of transmembrane segments in GmIPK2 protein using WaveTM server.

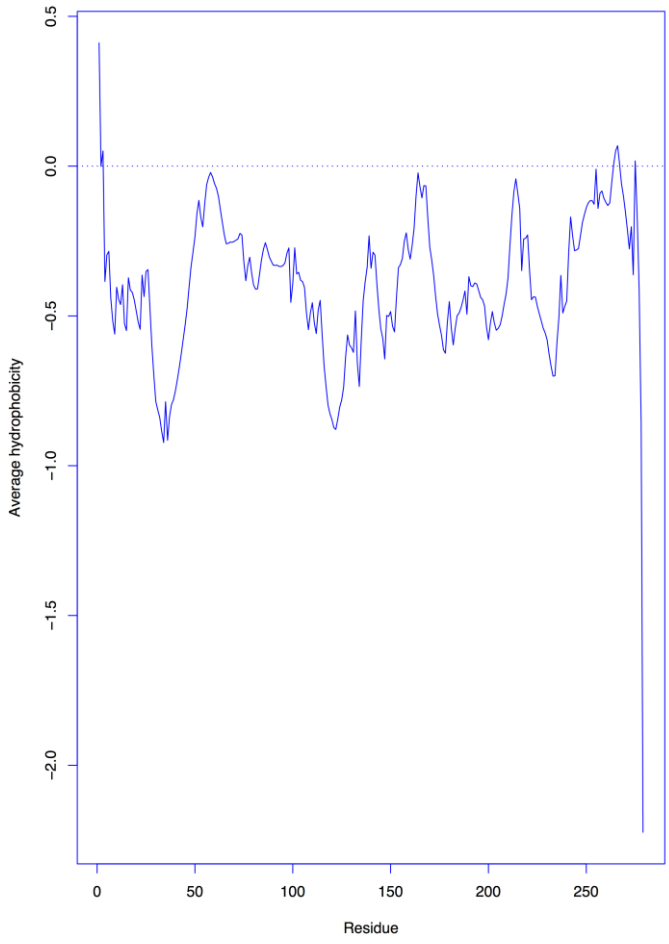


Figure 2: MEME suite motif analysis of selected IPK2 proteins showing eleven identified conserved ungapped motifs.

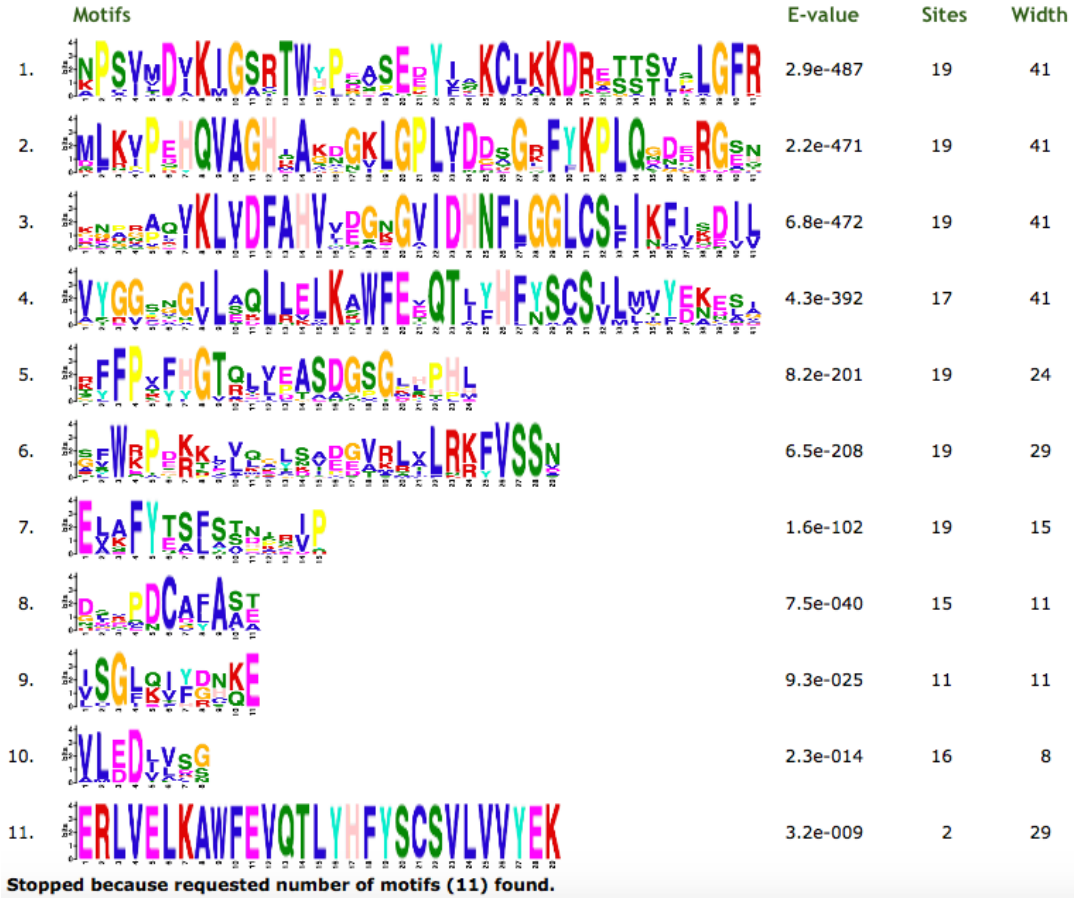


Figure 3: GmIPK2 protein disorder predicted using PrDOS server **(A)** Two-state prediction showing disordered residues in red at 5% false positive rate **(B)** Disorder probability plot of each residue with residues beyond the red threshold line predicted to be disordered.

(A)

1	**MLN**IPEHQVA	GHKAKDGILG	PLVDDFGKFY	KPLQTNKDDD	TRGSTELSFY	50
51	TSLAAAAHDY	SIRSFFPAFH	GTRLLDASDG	SGPHPHLVLE	DLLCGYSKPS	100
101	VMDVKIGSRT	WHLGDSEDYI	CKCLKKDRES	SSLPLGFKIP	GVKDSISSWE	150
151	PTRKSLQCLS	AHGVALVLNK	FVSSNNINHD	DHHPDCAFAT	EVYGAVLERL	200
201	QKLKDWFEVQ	TVYHFYSCSV	LVVYEKDLGE	RKATNPLVKL	VDFAHVVDGN	250
251	GVIDHNFLGG	LCSFIKFLKD	ILAV**ACLHK**			300

(B)

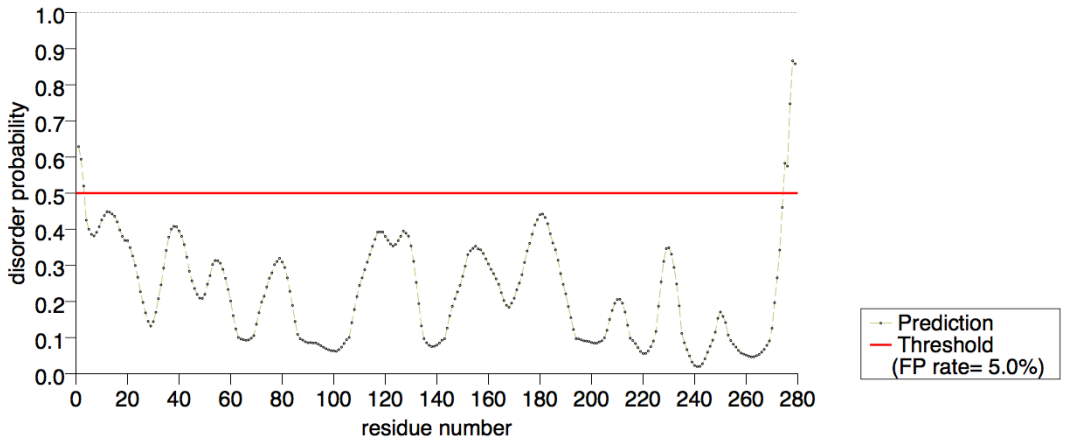
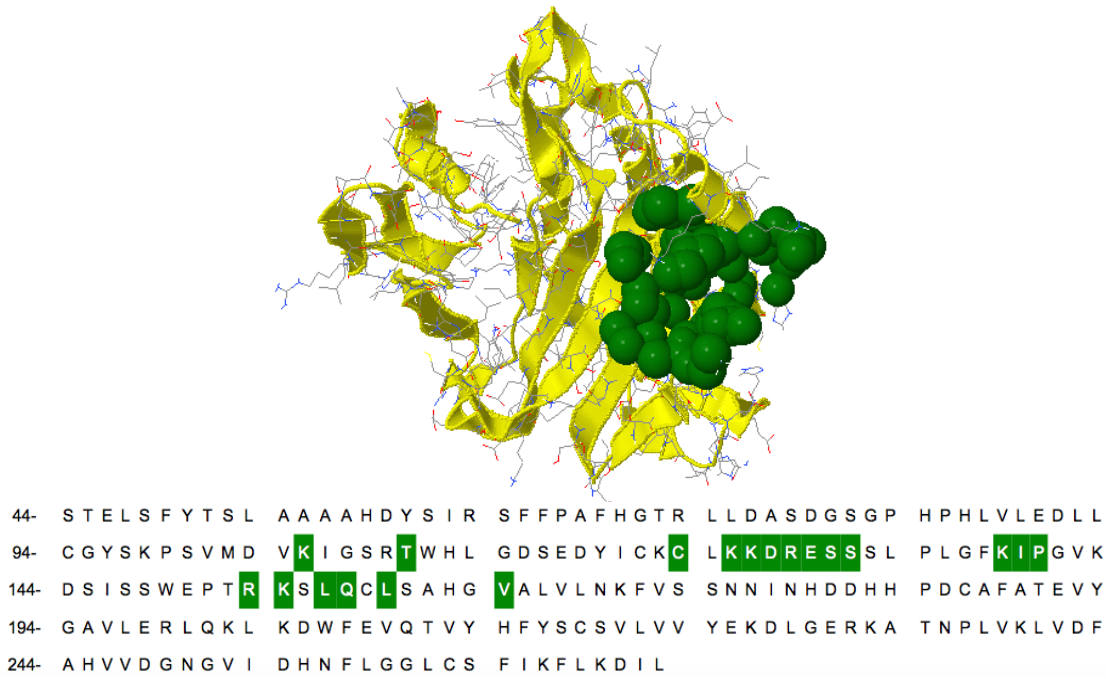
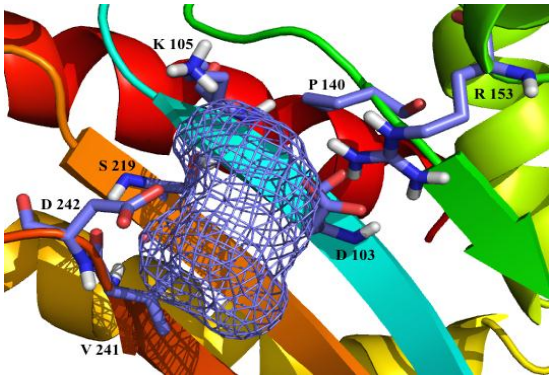


Figure 4: GmIPK2 ligand binding site predicted by online servers **(A)** CASTp (in green) **(B)** FTSite (in purple) **(C)** FunFOLD2 (in grey). The most likely ligand binding residues are marked in each model.

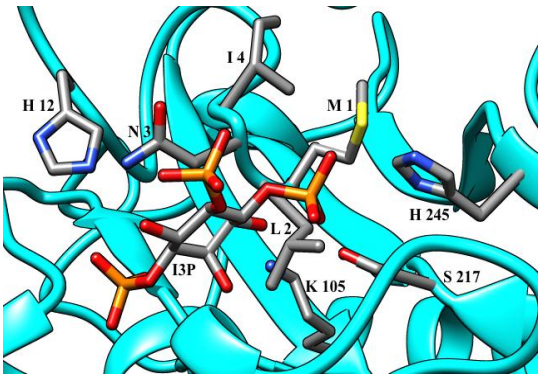
(A)



(B)



(C)



Parameters for molecular dynamics simulation using Gromacs 4.5.5

Parameters for protein (GmIPK2) MD simulation in water

```
; RUN PARAMETERS
integrator      = md        ; leap-frog integrator
nsteps          = 50000000  ; 2 * 25000000 = 50000 ps, 50 ns
dt              = 0.002     ; 2 fs

; OUTPUT CONTROL
nstxout         = 1000      ; save coordinates every 2 ps
nstvout         = 1000      ; save velocities every 2 ps
nstxtcout       = 1000      ; xtc compressed trajectory output every 2 ps
nstenergy       = 1000      ; save energies every 2 ps
nstlog          = 1000      ; update log file every 2 ps

; BOND PARAMETERS
continuation    = yes       ; Restarting after NPT
constraint_algorithm = lincs ; holonomic constraints
constraints     = all-bonds ; all bonds (even heavy atom-H bonds) constrained
lincs_iter      = 1         ; accuracy of LINCS
lincs_order     = 4         ; also related to accuracy

; NEIGHBORSEARCHING
ns_type         = grid      ; search neighboring grid cells
nstlist         = 5         ; 10 fs
rlist           = 1.0       ; short-range neighborlist cutoff (in nm)
rcoulomb        = 1.0       ; short-range electrostatic cutoff (in nm)
rvdw            = 1.0       ; short-range van der Waals cutoff (in nm)

; ELECTROSTATICS
coulombtype     = PME       ; Particle Mesh Ewald for long-range electrostatics
pme_order       = 4         ; cubic interpolation
```

```
fourierspacing = 0.16          ; grid spacing for FFT

; TEMPERATURE COUPLING IS ON
tcoupl         = V-rescale     ; modified Berendsen thermostat
tc-grps        = Protein Non-Protein          ; two coupling groups - more accurate
tau_t          = 0.1   0.1     ; time constant, in ps
ref_t          = 300   300     ; reference temperature, one for each group, in K

; PRESSURE COUPLING IS ON
pcoupl         = Parrinello-Rahman  ; Pressure coupling on in NPT
pcoupltype     = isotropic     ; uniform scaling of box vectors
tau_p          = 2.0           ; time constant, in ps
ref_p          = 1.0           ; reference pressure, in bar
compressibility = 4.5e-5       ; isothermal compressibility of water, bar^-1

; PERIODIC BOUNDARY CONDITIONS
pbc            = xyz           ; 3-D PBC

; DISPERSION CORRECTION
DispCorr       = EnerPres      ; account for cut-off vdW scheme
```

Parameters for protein-ligand (GmIPK2-I3P and GmIPK2-I0P) complex MD simulation

```
; RUN PARAMETERS
integrator     = md            ; leap-frog integrator
nsteps         = 25000000      ; 2 * 25000000 = 50000 ps, 50 ns
dt             = 0.001         ; 1 fs

; OUTPUT CONTROL
nstxout        = 0       ; suppress .trr output
nstvout        = 0       ; suppress .trr output
```

```
nstenergy       = 5000   ; save energies every 10 ps
nstlog          = 5000     ; update log file every 10 ps
nstxtcout       = 5000     ; write .xtc trajectory every 10 ps
energygrps      = Protein I3P / Protein I0P

; BOND PARAMETERS
continuation            = yes           ; continuing from NPT
constraint_algorithm = lincs          ; holonomic constraints
constraints             = h-bonds       ; bonds to H are constrained
lincs_iter              = 1             ; accuracy of LINCS
lincs_order             = 4             ; also related to accuracy

; NEIGHBORSEARCHING
ns_type                 = grid          ; search neighboring grid cells
nstlist                 = 5             ; 10 fs
rlist                   = 1.4           ; short-range neighborlist cutoff (in nm)
rcoulomb                = 1.4           ; short-range electrostatic cutoff (in nm)
rvdw                    = 1.4           ; short-range van der Waals cutoff (in nm)

; ELECTROSTATICS
coulombtype             = PME           ; Particle Mesh Ewald for long-range
electrostatics
rcoulomb                = 1.2
pme_order               = 4             ; cubic interpolation
fourierspacing          = 0.16          ; grid spacing for FFT

; TEMPERATURE COUPLING
tcoupl                  = V-rescale             ; modified Berendsen thermostat
tc-grps                 = Protein_I3P  Water_ions  ; two coupling groups-more accurate
tau_t                   = 0.1      0.1          ; time constant, in p
```

```
ref_t                = 300     300                ; reference temperature, one for each
group, in K

; PRESSURE COUPLING
pcoupl          = Parrinello-Rahman  ; Pressure coupling on in NPT
pcoupltype       = isotropic    ; uniform scaling of box vectors
tau_p            = 2.0          ; time constant, in ps
ref_p            = 1.0          ; reference pressure, in bar
compressibility = 4.5e-5        ; isothermal compressibility of water, bar^-1

; PERIODIC BOUNDARY CONDITIONS
pbc              = xyz          ; 3-D PBC

; DISPERSION CORRECTION
DispCorr         = EnerPres     ; account for cut-off vdW scheme;
```

Appendix II

Competent cell preparation

Competent cells of *E. coli* strain DH5α were prepared following classical calcium chloride method described by Sambrook and Russel, (2001). A single *E. coli* colony was picked from a freshly streaked LB plate and inoculated into 5 ml LB broth. The inoculum was incubated overnight at 37^{o}C in a shaker incubator at 200 rpm. 1 ml of saturated overnight culture was inoculated into 100 ml LB broth and shaken at 37^{o}C and 200 rpm until a OD_{600} of 0.3-0.4 was achieved. The culture was then transferred to pre-chilled, sterile centrifuge tubes and placed on ice for 10 min. The cells were harvested by centrifugation at 5000 rpm for 10 min at 4^{o}C, the supernatant was decanted and the pellet was gently resuspended in 10 ml ice cold 0.1M $MgCl_2$ solution. The suspension was centrifuged at 5000 rpm for 10 min at 4^{o}C and the pellet was resuspended in 10 ml ice cold 0.1M $CaCl_2$ solution. The suspension was then placed on ice for 30 min, following which the cells were recovered by centrifugation at 5000 rpm for 10 min at 4^{o}C. The pellet was gently resuspended in 1ml of ice cold 50mM $CaCl_2$/15% Glycerol. Aliquots of 100 µl were dispensed in microfuge tubes, flash frozen in liquid nitrogen and immediately transferred to -80^{o}C for storage.

Transformation of competent cells

For transformation, 100 µl competent *E. coli* DH5α cells were first briefly thawed on ice. 10 µl of the ligation reaction was then added to the competent cells, mixed by

gentle tapping and incubated on ice for 30 min. A no plasmid control tube was also prepared to check the efficiency of competent cells. The cells were then given a heat shock at 42^{o}C for exactly 2 min and were immediately chilled on ice. 900 µl of LB was added to each tube and incubated at 37^{o}C, 200 rpm in a shaker incubator for 1 hr. The tubes were centrifuged at 5000 rpm for 5 min and the pelleted cells were resuspended in ~100 µl of supernatant. The obtained suspension was aseptically plated on LA containing appropriate selection (50µg/ml ampicillin, 2% X-gal and 0.1M IPTG in case of blue-white screening) and incubated overnight at 37^{o}C. White colonies were selected randomly from the plate and streaked onto a master plate for further analysis.

Plasmid isolation

All plasmid DNA isolations were performed by following the alkaline lysis method described by Sambrook and Russel, (2001). A single colony of interest was picked from a freshly streaked LB plate and inoculated into 5 ml LB broth containing ampicillin (100 µg/ml) or kanamycin (50 µg/ml) and incubated overnight at 37^{o}C in a shaker incubator at 200 rpm. 1.5 ml of overnight grown culture was dispensed in microfuge tubes and pelleted by centrifugation at 10,000 rpm for 2 min. The supernatant was discarded and the pellet was resuspended in 100 µl of cell suspension solution by pipetting several times. The tubes were incubated on ice for 5 min following which 200 µl of freshly prepared lysis buffer was added and the contents of the tubes were mixed by inverting gently. 200 µl of neutralization buffer was finally added and the contents were mixed once again by gently inverting. The tubes were then centrifuged at 12,000 rpm for 15 min. The clear supernatant was transferred to fresh tubes and the plasmid DNA was precipitated by adding 480 µl of isopropanol,

incubated at -20°C for 10-15 min. The plasmid DNA was recovered by centrifugation at 12,000 rpm for 10 min and washed twice with 70% ethanol. The tubes were then centrifuged at 12,000 rpm for 5 min, the supernatant was discarded and any traces were removed carefully by aspiration. The pellets were allowed to air dry to remove residual alcohol by evaporation. Once all the alcohol was evaporated, the pellets were resuspended in 50 µl TE buffer. 20 µl of isolated plasmid was finally treated with 2 µl of 10 mg/ml RNaseA by incubation at 37^{o}C for 30 min and re-extracted by adding equal volume of phenol:chloroform-isoamyl alcohol mixture (1:1). The upper aqueous phase was collected and the plasmid was precipitated by adding 1/10 volume of neutralization buffer. The plasmid DNA was washed with 70% ethanol and dried following the procedure described above and resuspended in 20 µl TE buffer.

Solutions and Buffers

Competent cell

0.1M $MgCl_2$ — Dissolve 101.65 g of $MgCl_2.6H_2O$ in 450 ml of DDW and make up the volume to 500 ml. Sterilize the solution by autoclaving at 121^{o}C and 15 psi for 15 min.

0.1M $CaCl_2$ — Dissolve 2.7 g of $CaCl_2.2H_2O$ in 80 ml of DDW and make up the volume to 100 ml. Sterilize the solution by autoclaving at 121^{o}C and 15 psi for 15 min.

0.1M IPTG — Dissolve 0.238 g of IPTG in 10 ml water. Sterilize the pre-pared stock by passing through a 0.22 µm syringe filter and store in 1 ml aliquots at -20^{o}C for up to 1 year.

2% X-Gal (w/v)

Dissolve 200 mg of X-Gal in 10 ml dimethyl-formamide (DMF). Store 1 ml aliquots at -20^{o}C in amber microfuge tubes to protect from light.

Plasmid isolation

Cell Suspension Solution

25mM Tris (pH 8.0)
10mM EDTA
50mM Glucose

Lysis Buffer

0.2N NaOH
2.0% SDS

Neutralization Buffer

3M Sodium acetate (pH 5.2). Adjust the pH using glacial acetic acid.

RNase A (10 mg/ml)

Dissolve 100 mg of RNase A in 10 ml of 10 mM Tris-Cl (pH 7.5)/15 mM NaCl solution. Heat to 100^{o}C for 5 min and cool at room temperature. Store in 500 µl aliquots at -20^{o}C.

Electrophoresis

5X TBE (1 L)

54 g Tris base
27.5 g Boric acid
20 ml of 0.5 M EDTA (pH 8.0)

0.5M EDTA (pH 8.0)

Dissolve 186.1 g of EDTA in 800 ml DDW and stir vigorously on a magnetic stirrer. Adjust the pH by adding NaOH pellets. Make up the volume to 1 L and sterilize by autoclaving at 121^{o}C and 15 psi for 20 min.

6X DNA Loading Dye	10 mM Tris-HCl (pH 7.6) 0.03% Bromophenol blue 0.03% Xylene cyanol FF 60% Glycerol 60 mM EDTA
2X RNA Loading Dye	95% Formamide 0.025% SDS 0.025% Bromophenol blue 0.025% Xylene cyanol FF 0.5 mM EDTA
EtBr (10 mg/ml)	Dissolve 100 mg EtBr in 8 ml DDW and stir using a tube rotator to mix all the contents. Make up the volume to 10 ml and store at 4oC in amber vials to protect from light.
1% Denaturing Agarose Gel	Add 0.5 g agarose to 50 ml 0.5X TBE and boil. Cool the gel down to around 60^{o}C and add 29 µl 1M GTC and 1 µl EtBr. Swirl the solution and pour into a casting tray gently to avoid any bubbles. Allow the gel to solidify.

Southern Blotting

Depurination (0.25M HCl)	Add 20.7 ml of HCl (37%) to 800 ml DDW and make up the volume to 1 litre. Autoclave the solution at 121^{o}C, 15 psi for 20 min and store at RT.
Depurination (0.25M HCl)	Add 20.7 ml of HCl (37%) to 800 ml DDW and

	make up the volume to 1 litre. Autoclave the solution at 121^{o}C, 15 psi for 20 min and store at RT.
Denaturation (1 L)	88 g NaCl 20 g NaOH Autoclave at 121^{o}C, 15 psi for 20 min and store at RT.
Neutralization (1 L, pH 7.0)	88 g NaCl 60.6 g Tris-HCl (pH 7.2) Autoclave at 121^{o}C, 15 psi for 20 min and store at RT.
20X SSC (1 L, pH 7.0)	175.3 g NaCl 88.2 g Sodium citrate Autoclave at 121^{o}C, 15 psi for 20 min and store at RT.
100X Denhardt's (50 ml)	2% w/v BSA 2% w/v Ficoll 400 2% w/v PVP Sterilize using a 0.22 μm syringe filter and store at -20^{o}C.
0.1M Phosphate Buffer (pH 6.5)	Mix 68.5 ml of 0.2M monobasic sodium phosphate and 31.5 ml of 0.2M dibasic sodium phosphate. Make up the volume to 200 ml with DDW. Autoclave at 121oC, 15 psi for 15 min and store at RT.

Salmon Sperm DNA (10 mg/ml) Dissolve 50 mg ssDNA in 5 ml DDW sonicate to dissolve by giving 15s pulses. Sterilize using a 0.22 μm syringe filter. Denature the DNA before using by boiling for 5-10 min.

HPLC

2.4% HCl Add 5.67 ml HCl to 80 ml DDW and make up the volume to 100 ml.

2M HCl Add 17.13 ml HCl to 80 ml DDW and make up the volume to 100 ml.

5mM HCl Add 0.25 ml of 2M HCl to 80 ml DDW and make up the volume to 100 ml.

Pi assay

10% Ascorbic acid Dissolve 1 g of ascorbic acid in 10 ml DDW. Store at 4^{o}C for 1 week.

2.5% Ammonium molybdate Dissolve 1.25 g of ammonium molybdate in 50 ml DDW. Store at 4^{o}C for 1 month.

3M H_2SO_4 Slowly add 8.26 ml of concentrated H_2SO_4 to 30 ml DDW and make up the volume to 50 ml.

25% TCA Dissolve 12.5 g of TCA in 20 ml DDW and make up the volume to 50 ml. Store at 4^{o}C for 2 years.

50mM $MgCl_2$ Add 25 ml of 0.1M $MgCl_2$ to 25 ml DDW.

Bioavailability assay

α-amylase solution (pH 7.0)	12500 units/l α-amylase
	1.5 g/l NaCl
	1.5 g/l K_2HPO_4
	0.5 g/l Na_2CO_3
Stomach medium (pH 4.0)	0.1 g/l lipase
	0.125 g/l pepsin
	3.1 g/l NaCl
	1.1 g/l KCl
	0.6 g/l Na_2CO_3
	0.11 g/l $CaCl_2$
2% pancreatic solution	20 g/l pancreatin
	5.0 g/l bile
	5.0 g/l NaCl
	0.68 g/l KH_2PO_4
	0.3 g/l Na_2HPO_4
	0.84 g/l $NaHCO_3$

Plant Culture Media Composition

Components (mg/l)	B5	MS
Macronutrients		
KNO_3	2500.0	1900.0
$CaCl_2.2H_2O$	150.0	440.0
$(NH_4)_2SO_4$	134.0	
$MgSO_4.7H_2O$	250.0	370.0
$NaH_2PO_4.H_2O$	150.0	
NH_4NO_3		1650.0
KH_2PO_4		170.0
Micronutrients		
H_3BO_4	3.0	6.2
$MnSO_4.H_2O$	10.0	16.9
$ZnSO_4.7H_2O$	2.0	8.6

$CoCl_2.6H_2O$	0.025	0.025
$Na_2MoO_4.2H_2O$	0.25	0.25
$CuSO_4.5H_2O$	0.025	0.025
KI	0.75	0.83
Na_2EDTA	37.3	37.3
$FeSO_4.7H_2O$	27.8	27.8
Vitamins and other supplements		
Thiamine HCl	0.1	10.0
Pyridoxine HCl	0.5	1.0
Nicotinic acid	0.5	1.0
Inositol	100.0	100.0
Glycine	2.0	2.0

Plant Growth Regulators

BAP (1 mg/ml)

Add 1N NaOH dropwise to 10 mg BAP until it dissolves completely. Make up the volume to 10 ml with sterile DDW and pass through a 0.22 μm syringe filter. Store in 1 ml aliquots at -20^{o}C.

IBA (1 mg/ml)

Dissolve 10 mg IBA in 1 ml of absolute ethanol. Make up the volume to 10 ml with sterile DDW and pass through a 0.22 μm syringe filter. Store in 1 ml aliquots at -20^{o}C.

GA3 (0.75 mg/ml)

Dissolve 7.5 mg IBA in 1 ml of absolute ethanol. Make up the volume to 10 ml with sterile DDW and pass through a 0.22 μm syringe filter. Store in 1 ml aliquots at -20^{o}C.

Other Plant Culture Stocks

100mM Acetosyringone

Dissolve 196 mg of acetosyringone in 10 ml di-methylsulfoxide (DMSO). Store in 1 ml aliquots.

1M DTT Dissolve 1.54 g of DTT in 10 ml sterile DDW and pass it through a 0.22 µm syringe filter. Store in 1 ml aliquots at -20^{o}C.

100mM Sodium Thiosulfate Dissolve 1.58 g of STS in 100 ml sterile DDW and pass it through a 0.22 µm syringe filter. Store in 1 ml aliquots at -20^{o}C.

1M L-Cysteine Dissolve 12.116 g of L-Cysteine in 100 ml DDW and pass it through a 0.22 µm syringe filter. Store in 1 ml aliquots at -20^{o}C in amber microfuge tubes to protect from light.

Zeatin (1 mg/ml) Dissolve 10 mg of zeatin in 10 ml 1N NaOH and pass it through a 0.22 µm syringe filter. Store in 1 ml aliquots at -20^{o}C.

Antibiotics

Carbenicillin (100 mg/ml) Dissolve 1 g carbenicillin in 10 ml sterile DDW and pass it through a 0.22 µm syringe filter. Store in 1 ml aliquots at -20^{o}C. Use at the working concentration of 100 mg/l.

Timentin (70 mg/ml) Dissolve 700 mg timentin in 10 ml sterile DDW and pass it through a 0.22 µm syringe filter. Store in 1 ml aliquots at -20^{o}C. Use at the working concentration of 70 mg/l.

Cefotaxime (200 mg/ml) Dissolve 2 g cefotaxime in 10 ml sterile DDW and pass it through a 0.22 µm syringe filter. Store in 1 ml aliquots at -20^{o}C.

Ampicillin (100 mg/ml) Dissolve 1 g ampicillin in 10 ml sterile DDW and pass it through a 0.22 μm syringe filter. Store in 1 ml aliquots at -20^{o}C. Use at the working concentration of 100 μg/ml.

Kanamycin (50 mg/ml) Dissolve 0.5 g kanamycin in 10 ml sterile DDW and pass it through a 0.22 μm syringe filter. Store in 1 ml aliquots at -20^{o}C. Use at the working concentration of 50 μg/ml.

Rifampicin (25 mg/ml) Dissolve 0.25 g rifampicin in 10 ml absolute methanol and pass it through a 0.22 μm syringe filter. Store in 1 ml aliquots at -20^{o}C. Resuspend before using at the working concentration of 25 μg/ml.

List of Publications & Conference Proceedings

Publications

1. **Mansi Punjabi,** Navneeta Bharadvaja, Monica Jolly, Anil Dahuja & Archana Sachdev. (2018). Development and Evaluation of Low Phytic Acid Soybean by siRNA Triggered Seed Specific Silencing of Inositol Polyphosphate 6-/3-/5-Kinase Gene. ***Frontiers in Plant Science.*** 9:804

2. **Mansi Punjabi,** Navneeta Bharadvaja, Archana Sachdev & Veda Krishnan. (2018). Molecular characterization, modeling and docking analysis of late phytic acid biosynthesis pathway gene, inositol polyphosphate 6-/3-/5-kinase, a potential candidate for developing low phytate crops. ***3 Biotech.*** 8:344.

Conferences

1. **Mansi Punjabi**, Amit Kumar Gupta, Alkesh Hada, Monica Jolly, Navneeta Bhardvaja & Archana Sachdev. (2014). Molecular characterization and construct designing of late phytic acid pathway gene encoding Inositol Polyphosphate 6-/3-/5- Kinase for its ectopic down regulation in developing soybean seeds to generate low phytate grains. **Poster presentation** in International Congress on Agriculture, Food Engineering and Environmental Sciences- Sustainable Approaches (AFEESSA-2014) at JNU, Delhi, India.

2. **Mansi Punjabi**, Nabaneeta Basak, Vanita Pandey, Monica Jolly & Archana Sachdev. (2014). *RNAi vector for speed specific down regulation of Inositol Polyphosphate 6-/3-/5-kinase expression to develop low phytate soybean with improved nutrient bioavailability*. **Oral presentation** in International Soybean Research Conference, SOYCON 2014 at Indore, India.

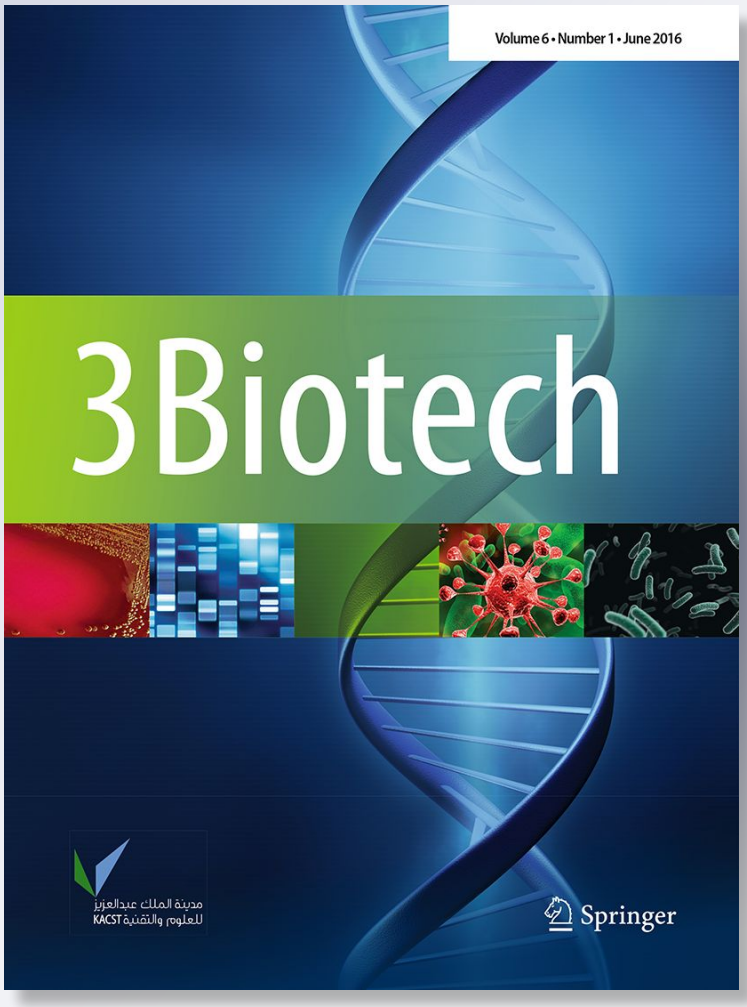
Molecular characterization, modeling, and docking analysis of late phytic acid biosynthesis pathway gene, inositol polyphosphate 6-/3-/5-kinase, *a potential candidate for developing low phytate crops*

Mansi Punjabi, Navneeta Bharadvaja, Archana Sachdev & Veda Krishnan

3 Biotech

ISSN 2190-572X
Volume 8
Number 8

3 Biotech (2018) 8:1-20
DOI 10.1007/s13205-018-1343-7



Springer

3 Biotech (2018) 8:344
https://doi.org/10.1007/s13205-018-1343-7

ORIGINAL ARTICLE

Molecular characterization, modeling, and docking analysis of late phytic acid biosynthesis pathway gene, *inositol polyphosphate 6-/3-/5-kinase*, a potential candidate for developing low phytate crops

Mansi Punjabi[1,2] · Navneeta Bharadvaja[1] · Archana Sachdev[2] · Veda Krishnan[2]

Received: 8 February 2018 / Accepted: 6 July 2018

Abstract

The coding sequence of *inositol polyphosphate 6-/3-/5-kinase* (*GmIPK2*) gene was identified and cloned from popular Indian soybean cultivar Pusa-16. The clone was predicted to encode 279 amino acids long, 30.97 kDa protein. Multiple sequence alignment revealed an inositol phosphate-binding motif, PxxxDxKxG throughout the IPK2 sequences along with other motifs unique to inositol phosphate kinase superfamily. Eight α-helices and eight β-strands in antiparallel β-sheets arrangement were predicted in the secondary structure of GmIPK2. The temporal analysis of *GmIPK2* revealed maximum expression in the seed tissues during later stages of development while spatially the transcript levels were lowest in leaf and stem tissues. Endosperm-specific cis-regulatory motifs (GCN4 and Skn_1) which support high levels of expression, as observed in the developing seeds, were detected in its promoter region. The protein structure of GmIPK2 was modeled based on the crystal structure of inositol polyphosphate multikinase from *Arabidopsis thaliana* (PDB:4FRF) and subsequently docked with inositol phosphate ligands (PDB: 5GUG-I3P and PDB: 4A69-I0P). Molecular dynamics (MD) simulation established the structural stability of both, modeled enzyme and ligand-bound complexes. Docking in combination with trajectory analysis for 50 ns MD run confirmed the participation of Lys105, Lys126 and Arg153 residues in the formation of a network of hydrogen bonds to stabilize the ligand-receptor interaction. Results of the present study thus provide valuable information on structural and functional aspects of *GmIPK2* which shall assist in strategizing our long-term goal of achieving phytic acid reduction in soybean by genetic modification of its biosynthetic pathway to develop a nutritionally enhanced crop in the future.

Keywords Glycine max · Phytic acid · Inositol polyphosphate 6-/3-/5-kinase (IPK2) · Low phytate crops, spatiotemporal expression · Homology modeling · Molecular docking · Molecular dynamics simulation

Introduction

Soybean is a phenomenal crop with a unique nutrient profile. It is the only plant source which provides a complete protein with quality equivalent to animal protein yet its human consumption is found to be very limited, even in the countries where its production is very high. Besides, majority of its uses in the food industry owe primarily to its functional properties rather than nutritional gain. The above stated can be attributed to the presence of a variety of bioactive components including phytic acid (PA), saponins, protease inhibitors, isoflavones, lectins, etc. Amongst these, PA is probably the most well-known antinutrient that is found in abundance in cereal grains and legumes. It is a saturated cyclic acid whose metabolism has been substantiated to regulate phosphorous and myo-inositol homeostasis. Due to its highly negative chemical structure, it forms phytate–mineral–protein complexes which result in reduced bioavailability of phosphorus, other associated minerals and proteins and thus contribute to the debatable health effects associated with its consumption. Therefore, in a need to reduce the level of this compound several approaches were undertaken which blocked its constitutive

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s13205-018-1343-7) contains supplementary material, which is available to authorized users.

✉ Archana Sachdev
arcs_bio@yahoo.com

[1] Department of Biotechnology, Delhi Technological University (Formerly Delhi College of Engineering), New Delhi 110042, India

[2] Division of Biochemistry, Indian Agricultural Research Institute, New Delhi 110012, India

Published online: 28 July 2018

synthesis, but often resulted in crop with deplorable agronomic performance (Bilyeu et al. 2008; Cichy and Raboy 2009; Feng and Yoshida 2004; Kuwano et al. 2009; Nunes et al. 2006). Recently trending reverse genetic approaches that target tissue-specific knockdown provide a promising alternative to eliminate this negative impact (Ali et al. 2013a, b). Research is thus underway to characterize PA biosynthetic pathway enzymes (Josefsen et al. 2007; Krishnan et al. 2015; Stiles et al. 2008; Sun et al. 2007; Sweetman et al. 2006, 2007) and study their expression patterns (Bhati et al. 2014; Fileppi et al. 2010; Suzuki et al. 2007) to generate adequate information for developing *lpa* soybean through metabolic engineering.

PA biosynthesis follows a succession of phosphorylation steps (Brearley and Hanke 1996a, b). One of its late pathway enzyme inositol polyphosphate 6-/3-/5-kinase (IPK2, EC 2.7.1.151) is a multiple specificity enzyme which phosphorylates the D-6, D-3, and/or D-5 positions on a variety of inositol polyphosphate substrates (InsP3/InsP4/InsP5) (Frederick et. al. 2005) that makes it a key enzyme in regulating PA turnover and is thus a crucial target for perturbing PA dynamics (Stevenson-Paulik et al. 2005). The enzyme has previously been studied in *Arabidopsis* (Stevenson-Paulik et al. 2002), yeast (Saiardi et al. 2000; York et al. 1999), mouse (Frederick et al. 2005) and human (Majerus 1992, 1996). A study has also been reported in soybean (Stiles 2007), however, it discusses only the gene's expression and enzyme kinetics. Therefore, with the intention to extend the current knowledge on its transcript expression (both spatial and temporal) as well as to understand its structure, biochemical features and evolutionary history, we investigated this gene in soybean and compared it with those of another 19 homologs in plants through various bio-computational tools. Furthermore, to elucidate complete molecular and biochemical mechanisms regulating PA synthesis as well as inositol metabolism and signalling in plants, an insight into IPK2 protein's structural features is required. Very less information regarding the three-dimensional structure of plant IPK2 is available. To facilitate the same, we determined the three-dimensional (3D) model of soybean GmIPK2 protein through homology modeling and performed its docking simulation with the substrate [1D-*myo*-inositol 1,4,5-trisphosphate (PDB: 5GUG-I3P) and 1D-*myo*-inositol-1,4,5,6-tetra*kis*phosphate (PBD: 4A69-I0P)] molecules.

Through this work, we made an effort to study *IPK2* gene to lay a groundwork for achieving the main goal of our work which is, to develop low phytic acid soybean, a nutritionally and agriculturally important crop.

Materials and methods

Plant material

Field grown soybeans (Glycine max [L.] Merr. cv. Pusa-16) were procured from Division of Genetics, IARI, New Delhi for use in this study. For spatial and temporal expression profiling, root, stem, leaf, and flower tissues were collected from 30-day-old plants while developing seeds were collected regularly after flowering until maturation and sorted based on their sizes. After collection, the tissue samples were immediately frozen in liquid nitrogen and stored at − 80 °C until used.

PCR amplification, cloning and sequencing of partial *GmIPK2* sequence

Total RNA was isolated from soybean seeds 8 mm in size using TRIzol reagent method (Invitrogen, USA). Approximately, 1 µg of isolated total RNA was used for single stranded cDNA synthesis by reverse transcription with RevertAid™ H Minus First Strand cDNA synthesis kit (ThermoFisher Scientific, USA). Prior to cDNA synthesis, the RNA was treated with RNase-free DNase I (Thermo Scientific, USA) for removal of any residual genomic DNA. Polymerase chain reaction (PCR) amplification was performed using 0.2 µg of the initial RT reaction optimized at 94 °C for 4 min; 35 cycles of 94 °C for 30 s, 62.5 °C for 30 s, 72 °C for 30 s; 72 °C for 10 min, using the *GmIPK2* specific primers: *IPK2*F 5′-ATGCTCAAGATC CCGGAG-3′ and *IPK2*R 5′-CAGTTAGTCTGCGACACT AATTCAAGC-3′. On completion, 2 µl of amplified product was analyzed by electrophoresis on 1% agarose gel and purified using gel/PCR DNA Fragments Extraction Kit (DF100) by Geneaid. The purified amplicon was subsequently ligated in pGEM®-T Easy vector (Promega, USA) by TA cloning following protocol described in the manual and used directly for transformation of competent *Escherichia coli* DH5α cells. The recombinant plasmids were identified by blue-white screening and confirmed by restriction digestion with *EcoRI* to release the amplicon fragment. The nucleotide sequences of *GmIPK2* thus isolated was determined by DNA sequence analysis of overlapping plasmid clones using universal primers (SP6 and T7) on an automated sequencer (ABI 3730xl DNA Analyzer, USA). The nucleotide sequence data was submitted to INSDC database GenBank.

Gene expression analysis by semi-quantitative reverse transcription PCR and quantitative real-time PCR

To analyze *GmIPK2* gene expression in different soybean tissues, we first performed semi-quantitative reverse transcription PCR (RT-PCR) to obtain an expression pattern

and then further estimated the transcript levels by quantitative real-time PCR (qRT-PCR). For RT-PCR, first strand cDNA was synthesized from total RNA using RevertAid™ H Minus First strand cDNA Synthesis Kit (Thermo Scientific, USA) and PCR was subsequently performed using the same pair of *GmIPK2* specific primers and thermal cycling conditions which were described for its cloning. We further monitored the quantitative amplification of *GmIPK2* by performing qRT-PCR analysis on a PikoReal 96 Real-Time PCR platform (ThermoFisher Scientific, USA). The reactions were set up using DyNAmo Flash SYBR Green qPCR Kit (Thermo Scientific, USA) with cDNA first strands as the template DNA. The expression of *GmIPK2* was normalized to an endogenous control, the housekeeping gene phosphoenolpyruvate carboxylase (*PEPCo*) (Sugimoto et al. 1992; Tuteja et al. 2004). The primers for the gene (q*IPK2*F 5′-CGCGGATCCGCGTTGCAGAAGCTCAAG-3′ and q*IPK2*R 5′-TCCCCGCGGGGAGCGACACTAATT CAAG-3′) and the internal control (q*PEPCo*F 5′-CATGCA CCAAAGGGTGTTTT-3′ and q*PEPCo*R 5′-TTTTGCGGC AGCTATCTCTC-3′) were designed using PrimerQuest tool by IDT, USA. The reactions were setup following the standard protocol provided in DyNAmo ColorFlash SYBR Green qPCR Kit (Thermo Scientific). The thermal profile used for PCR amplification was: 95 °C for 4 min; 40 cycles of 95 °C for 15 s, 60 °C for 30 s and fluorescence data collection. To minimize variation in the output, three technical replicates were carried out for each of the three biological replicates. The baseline data was collected for first 15 cycles to generate a baseline-subtracted plot of the logarithmic increase in fluorescence signal (ΔRn) versus cycle number. A standard fluorescence threshold (Rn) was set to 0.5 on the log fluorescence scale to determine the fractional cycle number (Ct value). The relative abundance of *GmIPK2* was calculated using the 2-ΔΔCT method (Livak and Schmittgen 2001). Dissociation curve analysis from 60 to 95 °C was also performed at the end of the assay to check for any non-specific amplification and/or contamination.

Sequence analysis and phylogenetic tree construction

Homologous *IPK2* sequences from other plants were identified with National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (Protein BLAST, https://blast.ncbi.nlm.nih.gov/Blast.cgi) using the above deduced *GmIPK2* sequence as a query. The primary sequence composition of IPK2 sequences were computed using PEPSTATS (Rice et al. 2000). The physio-chemical parameters of proteins were predicted using the ProtParam tool of ExPASy web server (Gasteiger et al. 2005). TargetP 1.1 Server using the cutoffs of 95% specificity was implemented for subcellular location prediction of GmIPK2 protein (Emanuelsson et al. 2000) and the outcome was compared to predictions obtained from MemType-2L (Chou and Shen 2007), WoLF PSORT (Horton et al. 2007), SubLoc v1.0 (Hua and Sun 2001) and CELLO v2.5 (Yu et al. 2006). Its transmembrane topology was predicted with PSIpred (Buchan et al. 2013), TMpred (Hofmann and Stoffel 1993) and NPS@ web programs (Rost et al. 1996). A final consensus was drawn manually and the topology was generated and visualized using Protter version 1.0 (Omasits et al. 2014). Presence of potential secretory signal peptides or mitochondrial targeting peptides was analyzed with SignalP 4.1 web server (Petersen et al. 2011). M-Coffee multiple sequence alignment (MSA) of the selected amino acid sequences was carried out to produce quality alignments which served as the basis for phylogenetic analysis (Wallace et al. 2006) to detect its evolutionary placement and phylogenetic similarity with other similar genes. The evolutionary tree was constructed using neighbor joining (NJ) clustering method to compute distances and poisson model for amino acid substitution in MEGA Version 6.0 (Molecular Evolutionary Genetic Analysis, Tamura et al. 2013). Bootstrap replications were set at 1000 to assess the degree of confidence for each clade of the observed tree. The final image was rendered with the Interactive Tree of Life server (iTOL) (Letunic and Bork 2016).

Promoter isolation and prediction of regulatory motifs

A motif search was carried out to define putative cis-elements in the promoter sequences involved in the regulation of *IPK2* expression using PlantCARE (Lescot et al. 2002). To identify these cis-regulatory elements, around 2 kb upstream sequence of the *IPK2* homologs were retrieved by NCBI's nucleotide BLAST program and fed to PlantCARE web tool. The tool's database identified regulatory elements in the isolated upstream sequences and each of the elements observed were analyzed with previously reported properties of the particular element. Unlike CDS, the regulatory sequences indirectly influence their immediate phenotype.

Secondary structure analysis and domain prediction

M-Coffee multiple sequence alignment (MSA) of the selected amino acid sequences was carried out and the alignment file was imported to Jalview (Waterhouse et al. 2009) to identify and shade the conserved amino acid sequences. Ungapped motifs were also detected using MEME web tool available on MEME suite 4.11.1 (Bailey et al. 2009). The motifs present were further verified using My Hits motif scan tool (Pagni et al. 2007). The domain composition was analyzed using CDD tool (http://www.ncbi.nlm.nih.gov/cdd) on NCBI server. A secondary structure consensus from

amino acid sequences was built based on the joint prediction with SOPMA (nearest-neighbor method) and PHD (neural networks method) correctly predicting 82.2% of the residues for 74% of co-predicted amino acids (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_seccons.html) (Geourjon and Deléage 1995; Rost and Sander 1993). Cysteine species and disulfide connectivity of protein sequences were determined using web tool DiANNA (Ferrè and Clote 2005). A secondary structure topology map of the 3D model was generated using ProMotif (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html) and rendered using TopDraw (Bond 2003).

Homology modeling and quality assessment of predicted model

The 3D model of GmIPK2 was constructed by template-based homology modeling using automated comparative protein modeling servers, SWISS-MODEL (Biasini et al. 2014) and PHYRE2 (Kelley et al. 2015) as well as a standalone comparative modeling program MODELLER 9.16 (Webb and Sali 2016). Comparative modeling consists of five main steps: search for related protein structures, selection of one or more appropriate templates, target-template alignment, model building and model evaluation. The final step of comparative modeling was additionally performed with RAMPAGE (Lovell et al. 2003), VERIFY 3D (Eisenberg et al. 1997) and ProSA servers (Wiederstein and Sippl 2007) to evaluate the stereochemical and energetic properties of the obtained models. In addition to this, the quality of models can also be assessed by structural comparison to the template using the MatchMaker tool in UCSF Chimera (Pettersen et al. 2004), a molecular visualization software package and calculating the Cα root mean square deviation scores (RMSDs) for each of the comparative models. The secondary structures of the final and template proteins were also compared by pairwise 3D alignment using MATRAS 1.2 (Kawabata 2003). Agreement on the best model was made on the basis of the majority in best scores from these different quality analyses.

Refinement of the predicted homology model

Molecular dynamics (MD) simulation was performed to optimize the obtained GmIPK2 model. Simulation of the model was conducted in explicit solvent using the GROMACS (Groningen Machine for Chemical Simulations) 4.5.5 package (Pronk et al. 2013). The model was solvated with simple point charge (SPC216) water in a cubic box with edges that were 0.7 nm from the molecular boundary. Initially, energy minimization (maximum number of steps: 1000) was performed to remove steric conflicts between the protein and water molecules, using the steepest descent integrator. The system was then equilibrated by optimizing the solvent molecules surrounding the energy-minimized model with NVT (constant Number of particles, Volume, and Temperature) and NPT (constant Number of particles, Pressure, and Temperature) ensembles using Berendsen thermostat and Parrinello-Rahman barostat, respectively. Finally, the system was simulated for 50 ns thrice maintaining the same temperature (300 K) and pressure (1 bar) using the Particle Mesh Ewald (PME) electrostatics method. After completion of production run, we collected the data using a post-processing tool, trjconv (strips out coordinates and correct for periodicity) and used this corrected trajectory as the input for studying the conformational stability of our simulated protein via *g_rms* and *g_rmsf* tools. We subsequently generated and analyzed the output plots using a simple plotting program, xmgrace (Turner 2005).

Active site prediction, molecular docking and MD simulation of the docked complexes

The refined protein model thus generated was docked to characterize the 3D structure of the complex, to gain an insight into the crucial amino acid residues also referred to as active residues that are involved in the complex formation. Based on the available literature for IPK2, 1D-*myo*-inositol-1,4,5-trisphosphate (PDB: 5GUG-I3P) and 1D-*myo*-inositol-1,4,5,6-tetra*kis*phosphate (PDB: 4A69-I0P) ligands were selected for optimized GmIPK2 protein receptor. Docking of these ligands was performed using AutoDock Vina (version 1.1.2) (Trott and Olson 2010) following the semi-flexible approach of docking. AutoDock Vina reads all the molecules in a simplified PDB file representation, termed PDBQT and thus the coordinates of both GmIPK2 protein and its ligands (downloaded from the RCSB protein data bank) were prepared using MGL tools (version 1.5.4) AutoDock tools (Morris et al. 2009) prior to docking. All the water and solvent atoms of the protein were removed and the polar hydrogen atoms were added. The GmIPK2 molecule was kept rigid while the ligands were allowed to rotate and explore more flexible binding pockets. Vina used a customized rectangular 3D cartesian grid for specifying the binding site of the protein and for efficient geometric scoring. The dimensions of the grid were customized to make sure that the size of the search space is large enough for the ligand to rotate in. Once set, the docking run produced ligand poses each with a definite binding energy (kcal/mol) calculated based on the scoring function used in Vina. The conformations with the lowest binding affinity were chosen and the interaction diagrams were generated using Discovery Studio Visualizer 4.1 (Accelrys Software Inc., USA 2013). The amino acid residues present at a distance of 2 Å were considered as the binding partners of the ligands. The active amino acid residues were also predicted by combining results of

three different interface prediction web servers, CASTp (Dundas et al. 2006), FTSite (Kozakov et al. 2015) and FunFOLD2 (Roche et al. 2013) into a consensus. The final complexes were equilibrated by MD simulation by following the procedure described for protein model in explicit water in the previous section but by applying restraints on the ligands to prevent them from moving away from the binding site. Once equilibrated, we simulated the complexes for 50 ns and analyzed the resulting trajectories for their stability using *g_rms, g_rmsf, g_hbond* and *g_mmpbsa* tools.

Results and discussion

Cloning and sequencing of *GmIPK2*

We first performed a BLASTN analysis in the plant comparative genomics portal Phytozome v9.1 (Goodstein et al. 2012) with the soybean *IPK2* gene sequence available in NCBI (GenBank: NM_001250522) to retrieve a 1241 bp transcript sequence. Based on the sequence data thus derived, we designed primers specific to amplify complete *GmIPK2* coding sequence (CDS) as well as a fragment of its 3′ untranslated region (UTR) with the aim to design silencing construct in the future specific to this highly conserved region because of its substantial role in gene regulation. To accomplish this, we converted total RNA isolated from 8 to 10 mm developing seed stage to cDNA and performed PCR amplification from the synthesized cDNA template following the protocol described under sect. "PCR amplification, cloning, and sequencing of partial GmIPK2 sequence". The amplicon was then cloned into pGEM-T easy vector system and introduced into the bacterial host *E. coli* (DH5α). The putative cDNA clones were verified by restriction analysis with EcoRI and sequence characterized to ~ 910 bp residues in length (Fig. 1). It contains a single open reading frame ~ 840 bp long which potentially encode a single polypeptide of 279 amino acid residues and a ~ 70 bp 3′ UTR fragment. We submitted the obtained CDS data to NCBI (GenBank: KF297702) and used the information as query to conduct protein homology search using PSI-BLAST algorithm. Amongst the sequences producing significant alignment, we shortlisted 30 plant IPK2 sequences based on percentage sequence identity to carry out further in silico analysis.

Spatial and temporal expression profiling

As discussed above, tissue-specific modulation of *IPK2* gene to generate a *lpa* mutant is essential to evade any possible pleiotropic effects. Before this can be achieved it is vital to investigate its spatial expression profile in different tissues as well as its temporal expression profile in developing seeds, to provide an initial point for strategic achievement of a desired level of silencing.

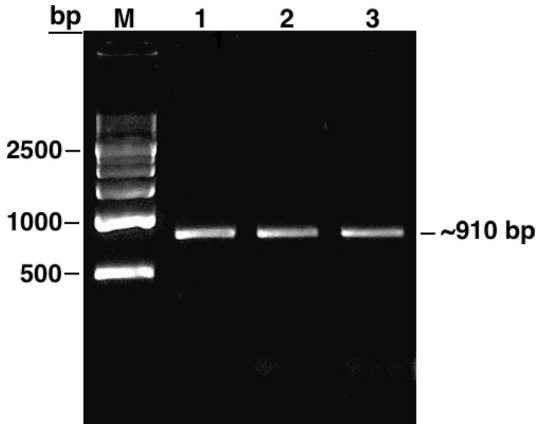


Fig. 1 Agarose gel showing PCR amplification product of *GmIPK2* gene isolated from developing seeds (8–10 mm) of *Glycine max cv.* Pusa-16. Lane M: 500 bp DNA ladder, lanes 1–3: ~ 910 bp *GmIPK2* fragment

To study the expression of *GmIPK2* gene in different tissues of a soybean plant and during seed development, semi-quantitative as well as real-time PCR expression analysis was performed with total RNA isolated from root, stem, leaf and flower tissues of 30-day-old *G. max* plants and developing seeds ranging from 0 to 16 mm in size distributed in eight different progressive stages. Semi-quantitative PCR analysis revealed a differential pattern of *GmIPK2* transcript expression across the set of experimental tissues analyzed, with the highest level of transcripts observed in seeds (Fig. 2a). The same was confirmed by steady-state qRT-PCR analysis which also detected highest level of transcripts in seeds (Fig. 2b). This suggests that *GmIPK2* play a key role in PA biosynthesis in this tissue for use as a primary source of energy during germination. In both the analyses, same level of amplification was observed for *PEPCo* housekeeping gene. Since PA is also required for several other vital functions throughout the plant system, a basal level of it is observed in all the tissues as well. Thus, amongst the other tissues analyzed, expression of *GmIPK2* was also recorded in roots and flowers, but at a level lower than that observed in the cotyledons. The strong presence of *GmIPK2* transcripts in these tissues may be attributed to its role in regulating cytosolic calcium gradient which correlates with pollen germination, pollen tube growth (Franklin-Tong et al. 1996; Malho 1998; Pierson et al. 1994; Xu et al. 2005), root growth, and root hair development (Bibikova et al. 1997; Felle and Hepler 1997; Wymer et al. 1997). A low level of expression was also observed in stems and leaves as it is studied to be involved in regulating a vital function of axillary shoot branching by participating in auxin signaling (Zhang et al. 2007).

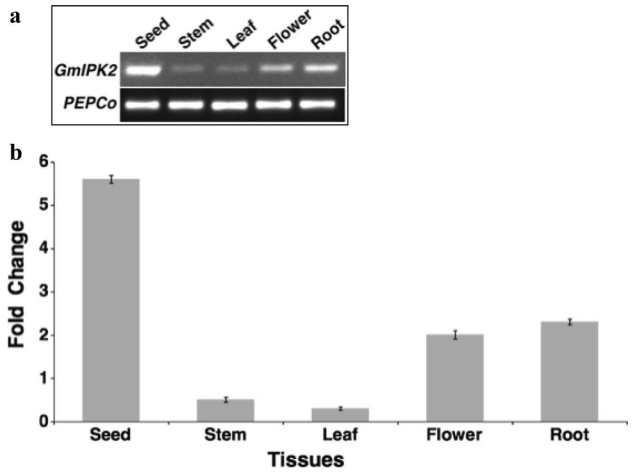


Fig. 2 **a** RT-PCR expression analysis of *GmIPK2* gene in different plant tissues of Pusa-16 cultivar using soybean housekeeping gene *PEPCo* as an internal control. **b** Relative quantification of *GmIPK2* transcript levels in the samples analyzed above by qRT-PCR, normalized to soybean housekeeping gene *PEPCo*. The leaf tissue was taken as calibrator. The data are mean of technical triplicates of each of the three biological replicates with error bars indicating standard deviation (SD)

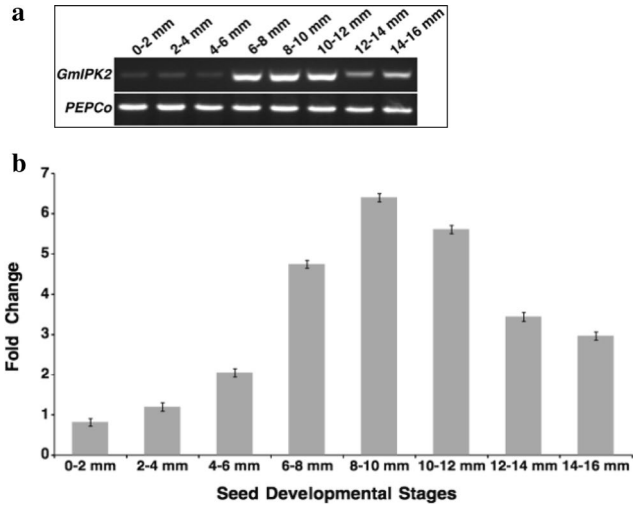


Fig. 3 **a** RT-PCR expression analysis of *GmIPK2* gene in developing seeds of Pusa-16 cultivar using soybean housekeeping gene *PEPCo* as an internal control. **b** Relative quantification of *GmIPK2* transcript levels in the samples analyzed above by qRT-PCR, normalized to soybean housekeeping gene *PEPCo*. The 0–2 mm seed stage was taken as calibrator. The data are mean of technical triplicates of each of the three biological replicates with error bars indicating SD

We then analyzed its temporal expression pattern during eight progressive seed development stages. In both semi-quantitative as well as real-time PCR expression analysis, we observed that its expression increased as the development progressed and reached a peak value at the later stages of seed development (Fig. 3). This pattern of expression coincides evenly with the pattern of accumulation of PA which is linear throughout most of seed development (Raboy and Dickinson 1987). Similar results were obtained in a microarray transcriptome study conducted in the past in our laboratory (GEO: GSE69821). This observation can be explained by the increase in production of phosphorous compounds required to support growth and development during initial stages of seed development when the synthesis of phosphorous reserve, PA is minimal (Raboy and Dickinson 1987).

The spatiotemporal analysis in summary identifies seed as the major tissue for its expression with maximum relative expression occurring during the later stages of its development. Thus, from the present study we can hypothesize that targeting the *GmIPK2* gene expression during late seed development stages may provide a potential strategy for generating *lpa* soybean with enhanced nutritional value.

Characterization of regulatory motifs in *IPK2* gene promoter

To understand the mechanism regulating the spatiotemporal expression pattern of *GmIPK2* gene thus observed, we analyzed the promoter region of all its homologs and identified different cis-regulatory elements located therein. A 2 kb sequence upstream to the open reading frame was identified and subjected to PlantCARE analysis. Database search revealed the presence of many motifs related to seed-specific promoters, hormone-responsive cis-elements (HRE) and cis-elements responsive to stresses (DSRE) that together contribute to the differential regulation of our gene (Table 1). Light responsive elements (Arguello-Astorga and Herrera-Estrella 1996; Feldbrugge et al. 1996; Lois et al. 1989; Lopez-Ochoa et al. 2007; Sessa et al. 1995) were observed most frequently which suggest a probable diurnal regulation of *IPK2* expression. The GCN4 and Skn_1 motifs highly conserved in the promoters of cereal seed storage protein genes were located within the *IPK2* promoter. These cis-acting elements play a central role in controlling endosperm-specific gene expression (Onodera et al. 2001; Takaiwa et al. 1996; Washida et al. 1999; Wu et al. 1998). Multiple HREs particularly those known to be involved in abscisic acid (ABA) and gibberellic acid (GA3) sensing were also identified. A ABA-responsive element (ABRE) with the core sequence PyACGTG/TC and three GA-responsive elements (GARE) AAACAGA, TCTGTTG and TATCCAC/T (Niu et al. 2016; Mongkolsiriwatana et al. 2009; Yamaguchi-Shinozaki et al. 1989) were identified which previously were reported to correlate with PA accumulation during grain filling (Abe et al. 2003; Matsuno and Fujimura 2014). Aggarwal et al. 2015 reported that *IPK2* is an ABA-induced gene which is antagonistically suppressed by GA3 underlining the crucial role played by these hormones in regulating PA pathway genes. Putative elements responsive

Table 1 Potential cis-acting elements identified in the 5′ regulatory sequences of plant *IPK2s*

Classification	Name	Sequence	Source organism	References
Endosperm	GCN_4 motif	GAAGCCA, TGAGTCA	*Oryza sativa*	Takaiwa et al. (1996), Onodera et al. (2001)
	Skn-1 motif	CAAGCCA, TGTGTCA, GTCAT	*Oryza sativa*	Washida et al. (1999)
ABA	ABRE	CACGTG, TACGTG	*Arabidopsis thaliana*	Yamaguchi-Shinozaki et al. (1989)
Gibberellin	GARE	AAACAGA, TCTGTTG, TATCCAC/T	*Brassica oleracea* *Oryza sativa*	Mongkolsiriwatana et al. (2009), Jun Niu et al. (2016)
Light	Box 4	ATTAAT	*Petroselinum crispum*	Lois et al. (1989)
	CATT-motif	GCATTC	*Zea mays*	Arguello-Astorga and Herrera-Estrella (1996)
	G-Box	CACGTG	*Pisum sativum, Arabidopsis thaliana, Zea mays*	Sessa et al. (1995), López-Ochoa et al. (2007)
	ATCC/T-motif	AATCTAATCC/T	*Pisum sativum, Arabidopsis thaliana*	Arguello-Astorga and Herrera-Estrella (1996)
	TCT-motif	TCTTAC	*Arabidopsis thaliana*	Arguello-Astorga and Herrera-Estrella (1996)
	Box-I	TTTCAAA	*Pisum sativum*	Arguello-Astorga and Herrera-Estrella (1996)
	ACE	AAAACGTTTA	*Petroselinum crispum*	Feldbrugge et al. (1996)
	GA-motif	AAGGAAGA	*Glycine max*	Arguello-Astorga and Herrera-Estrella 1996
MeJA	CGTCA-motif TGACG-motif	CGTCA TGACG	*Hordeum vulgare* *Hordeum vulgare*	Kim et al. (1993), Rouster et al. (1997)
Drought	MBS	T/CAACTG	*Arabidopsis thaliana*	Shinozaki and Yamaguchi-Shinozaki (2000), Abe et al. (2003)
Anaerobic	ARE	TGGTTT	*Zea mays*	Nguyen et al. (2003)

to methyl-jasmonate, drought inducibility and anaerobic induction were also observed which help combat assorted types of abiotic factors that plants are exposed to under natural environment (Abe et al. 2003; Kim et al. 1993; Nguyen et al. 2003; Rouster et al. 1997; Shinozaki and Yamaguchi-Shinozaki 2000).

Computation of physiochemical parameters and subcellular localization prediction

GmIPK2 sequence similarity search using PSI-BLAST revealed homology to IPK2 protein sequences from different plant sources showing maximum similarity with *Glycine soja* (98%) and *Vigna radiata* (74%) (Table 2). Primary sequence analysis of these homologs indicate that leucine is the most abundant amino acid which makes upto approximately 9–11 mol percent of its backbone residues whilst the percentage of tryptophan and methionine were found to be the least at approx. 1% across species. The isoelectric points (pI) of all IPK2 proteins were computed to be under 7 suggesting that they are most likely to precipitate in acidic buffers except *Glycine soja, Medicago truncatula, Zea mays and Sorghum bicolor* which has a pI of 7.1, 8.57, 10.48 and 8.26, respectively indicating their solubility in basic buffers. The calculated pI will be useful for empirical protein purification by isoelectric focusing and ion exchange chromatography. The extinction coefficient (EC) (Gill and Von Hippel 1989) of IPK2 proteins measured at 280 nm in water was found ranging from 25,440 to 34,380 M–1 cm–1 with respect to their concentration of aromatic amino acids (11–15%) and cystine (disulfide bonds). These EC values can be used to calculate protein concentration in a solution which in turn help in the quantitative study of biochemical interactions (protein–protein and protein–ligand). The instability indices (Ii) computed for selected IPK2 proteins are used to determine their in vivo half-lives (Guruprasad et al. 1990). Rogers et al., 1986 reported that proteins having Ii values greater than 40 have an in vivo half-life of less than 5 h while those proteins having Ii values less than 40 have a longer in vivo half-life of 16 h. Our study showed that Ii values of all the homologs are less than 40 and hence are thermally stable with a long half-life except for *Lotus japonicas* and *Zea mays* kinases that have a Ii above 40 which indicate their possible thermal instability. Thermostability of proteins result from a combination of several factors acting synergistically. Here we assess thermal stability of our proteins in direct proportionality with their aliphatic index (Ai) which is a measure of the relative volume occupied by aliphatic side chains (Ikai 1980). The Ai values determined for IPK2 kinases ranged from 68.79 to 96.70 with those from brassicaceae family showing lowest thermal stability. This in turn is indicative of their greater flexibility at a wide range of temperatures when compared to proteins of other families. Grand average of hydropathicity (GRAVY) number reflect

Table 2 Physiochemical parameters of shortlisted plant IPK2 sequences computed using the ProtParam tool

Organism	Accession no.	Sequence length	MW	pI	EC	Ii	Ai	GRAVY	−R	+R
Glycine max	AGW99177.1	279	30,979.2	6.29	29,910	29.72	89.07	−0.18	35	30
Glycine soja	KHN19419.1	178	19,898.87	7.10	23,950	29.73	95.73	−0.059	21	21
Vigna radiata	XP_022634636.1	286	31,482.10	6.59	28,420	38.17	93.67	0.045	29	27
Cajanus cajan	XP_020232596.1	267	29,463.71	6.86	28,420	33.70	91.31	−0.132	31	30
Phaseolus vulgaris	XP_007133857.1	264	28,951.0	6.59	26,930	38.90	90.38	−0.017	27	25
Lotus japonicus	AFK39224.1	283	31,348.76	5.66	32,430	42.36	87.74	−0.139	39	30
Cicer arietinum	XP_004510840.1	296	33,061.70	6.12	26,930	31.59	91.11	−0.250	38	33
Medicago truncatula	XP_003627882.1	426	47,961.8	8.57	29,910	31.89	85.02	−0.344	49	53
Arachis hypogaea	ALT56981.1	297	32,684.01	6.45	28,420	36.41	84.01	−0.227	34	32
Arachis duranensis	XP_015937330.1	297	32,883.25	6.45	28,420	37.35	84.01	−0.248	35	33
Corchorus capsularis	OMO94774.1	292	32,498.97	5.84	28,420	40.75	92.43	−0.159	38	33
Theobroma cacao	EOY22693.1	304	34,213.77	6.32	31,400	37.61	84.61	−0.292	39	36
Herrania umbratica	XP_021285810.1	305	34,392.01	6.56	36,900	40.38	83.38	−0.307	38	36
Durio zibethinus	XP_022738659.1	295	32,954.54	6.71	28,420	38.72	87.83	−0.237	36	35
Solanum lycopersicum	XP_004235863.1	294	32,476.9	5.93	28,420	34.32	83.23	−0.261	35	30
Solanum tuberosum	NP_001335929.1	408	44,295.0	5.96	28,420	20.21	93.19	−0.254	50	43
Brassica rapa	XP_009112000.1	274	30,640.7	6.25	34,380	30.25	80.66	−0.283	36	33
Capsicum baccatum	PHT32205.1	374	40,995.36	5.97	32,890	33.47	90.19	−0.297	50	44
Nicotiana attenuata	XP_019263239.1	368	40,865.55	5.51	31,400	30.30	93.51	−0.179	48	40
Capsicum annuum	PHT66064.1	374	41,015.41	5.91	32,890	32.27	92.01	−0.277	50	43
Capsicum chinense	PHU00941.1	374	40,986.37	5.81	32,890	32.27	92.27	−0.271	51	43
Brassica napus	XP_013749343.1	285	31,961.0	5.87	35,870	30.23	75.86	−0.387	40	34
Arabidopsis lyrata	XP_020870073.1	300	33,671.08	5.82	31,400	29.76	83.47	−0.324	42	35
Prunus avium	XP_021815221.1	281	31,100.28	6.20	31,400	26.98	86.41	−0.209	34	30
Trifolium pratense	PNY08557.1	273	30,275.83	5.50	23,950	29.99	96.70	−0.054	37	26
Arabidopsis thaliana	NP_200984.1	300	33,486.7	5.72	31,400	25.96	80.23	−0.329	40	31
Prunus persica	XP_007209445.1	282	31,303.57	6.50	31,400	28.39	85.39	−0.252	35	33
Lepidium latifolium	ACK86969.2	297	33,212.5	6.59	34,380	29.25	81.95	−0.285	31	29
Zea mays	XP_008649440.2	240	26,665.45	10.48	29,450	65.74	68.79	−0.517	22	37
Aegilops tauschii	XP_020147020.1	287	30,607.86	6.13	26,930	38.85	88.43	−0.052	33	29
Sorghum bicolor	XP_002452184.1	322	34,550.64	8.26	28,420	39.34	90.93	−0.013	32	34

MW molecular weight (g/mol), *pI* isoelectric point, *EC* extinction coefficient (M−1cm−1), *Ii* instability index, *Ai* aliphatic index, *GRAVY* grand average hydropathy, *(−R)* number of negative residues, *(+R)* number of positive residues

the average hydropathy of a protein, the positively rated being hydrophobic and negatively rated being hydrophilic in nature Kyte and Doolittle (1982). GRAVY index for IPK2 kinases was found ranging from −0.517 to −0.013 indicating that these proteins will interact favourably with water except for *Vigna radiata* with an index of 0.045 and hence is a potential hydrophobic protein.

Protein localization and target peptide predictions are significant studies as they aid in in silico protein function characterization as well as genome annotation. The acidic amino acid composition of IPK2 homologs determined by the physiochemical analysis conducted above suggest that they are cytoplasmic in nature as opposed to the basic amino acid composition of membrane proteins for their stability (Schwartz et al. 2001). Further sequence analysis based on TargetP scores (cTP: 0.163, mTP: 0.066, SP: 0.087, other: 0.906) also suggest that IPK2 kinases may be located anywhere in the cell besides chloroplast and mitochondria. The sequences were not predicted to have any signaling pre-sequence which was indicated by their low signal peptide (SP) score reinforced that they are soluble in nature. Besides, a consensus of predictions obtained from WoLF PSORT, CELLO v2.5, SubLoc v1.0 and MemType-2L servers (Table S1) also established the cytoplasmic character of IPK2 protein. However, TMpred, PSIPRED, NPS@ and DAS servers also identified a single consensus C-terminal transmembrane region positioned at 259–274. This could have been mistaken due to the presence of large regions

of hydrophobic residues in the C-terminal soluble region, as hydrophobicity alone is used as the criterion to predict membrane-spanning regions. The absence of any transmembrane helix is also well documented in the hydrophobicity plot generated using waveTM server (Fig S1) (Pashou et al. 2004) as majority of amino acids show negative for hydrophobicity.

Motif analysis and secondary structure characterization

Multiple sequence alignment (MSA) of the selected plant IPK2 homologs was performed using M-Coffee web server which is a meta-method for assembling MSA by combining the output of several individual methods into one to generate the best possible alignment. The consensus alignment thus generated revealed several significantly conserved motifs and sites unique to inositol phosphate kinases (Fig. 4). A signature inositol phosphate-binding motif, PxxxDxKxG was identified in all the aligned sequences (Odom et al. 2000; Saiardi et al. 1999) which confirms that they belong to inositol phosphate kinase (IPK) superfamily of IP kinases. Holmes and Jogl, 2006 state that the members of this superfamily share several strictly conserved signature motifs with each other and are predicted to assume the same overall fold, despite the low sequence conservation. The core catalytic tyrosine kinase motif, RxxxExxxY was also discovered in all the sequences which suggest that they are tyrosine-specific protein kinases (Cooper et al. 1984). IPK2 sequences from Solanaceae and Rosaceae families were found to contain a Glycine-rich consensus ATP-binding GxGxxG motif characteristic of protein kinase C (PKC) catalytic domain (Steinberg 2008). The classical PKC and plant CDPKs recognized phosphorylation S/TxK/R motif (Nishikawa et al. 1997; Neumann et al. 1996; Roberts and Harmon 1992) was also identified in some of the sequences speculating their role in lipid-dependent PA biosynthetic pathway. Such promiscuous kinase activity suggests that both lipid-dependent and independent pathways regulate PA biosynthesis as well as basic nuclear and cellular processes in plants (Josefsen et al. 2007; Stevenson-Paulik et al. 2002). A protein recognition LxxLL motif common to all of the aligned sequences indicate their participation in protein–protein interactions, regulating cell signalling, cell adhesion, and transcription (Plevin et al. 2005). Further analysis by MEME suite web server identified a total of 11 conserved ungapped motifs, with motif PxxxDxKxG being the most conserved amongst all IPK2 homologs as indicated by its lowest *E* value of 2.9e-487 (Fig S2). The obtained motifs were subjected to BLASTP analysis for conformation of their annotations which established that they all belong to IPK superfamily domain (CDD Acc: cl12283) and thus substantiate our previous results. We further explored GmIPK2 protein sequence by Motif Scan which recognized diverse protein kinase phosphorylation sites (Table S2). Since phosphorylation acts as a molecular switch in modulating protein function, structural rearrangement and cellular localization it can be suggested that GmIPK2 play a critical role in many biological regulatory events of signalling, proliferation, differentiation, and apoptosis. Its role in biological processes as diverse as mRNA export (York et al. 1999), DNA repair (Hanakahi et al. 2000), regulation of chromatin structure (Shen et al. 2003; Steger et al. 2003), maintenance of basal resistance to plant pathogens (Murphy et al. 2008) and apoptosis (Agarwal et al. 2009) have been studied in the past.

The secondary structure of a protein is more conserved than its nucleotide sequence and is, therefore, a prized source of information in understanding its classification, function, molecular evolution, and interaction with macromolecules (Reehana et al. 2013). In addition, secondary structure provides the first framework for homology based prediction of a protein 3D-model. Thus, in the current study, we inferred the secondary structure composition of IPK2 kinases from a three-state prediction done using NPS@ web server (Table 3). A high coil content was observed in most of the sequences including GmIPK2 while some showed them in nearly equal proportion with α-helix. This structural state can be justified based on the rich content of highly flexible glycine and kink inducing proline amino acid residues. The percentage of extended strands (% Ee) in all the kinases were found ranging from 13 to 29% except for the Poaceae family kinases which showed a low % Ee conformation (below 10%). PDBSum tool PROMOTIF analysis of GmIPK2 polypeptide identified total eight α-helices and eight β-strands arranged to form three antiparallel β-sheets, interspersed throughout by regions of coil or turn conformations (Fig. 5). We also recognized a varying number of bonded half-cystine pairs in all the IPK2 protein sequences using DIANNA server. It revealed the presence of 8 Cys residues in GmIPK2 and the most probable half-cystine pairs predicted by CYSREC were 94–121, 123–276 and 158–186. These potential long-term interactions participate in stabilizing the native conformations of our proteins and may as well contribute to differences in their tertiary structures.

Evolutionary analysis

Conserved motif analysis of IPK2 protein sequences point at a distinct evolutionary association between these kinases. Phylogenetic analysis would provide a further basis to determine their relatedness as well as to understand their collective evolution from a common ancestor. Previous research has reported that IPK superfamily of kinases to which *IPK2* belongs evolved from a common ancestor (Irvine and Schell 2001; Shears 2004). In our study, based on the alignment obtained in "Motif analysis and secondary structure

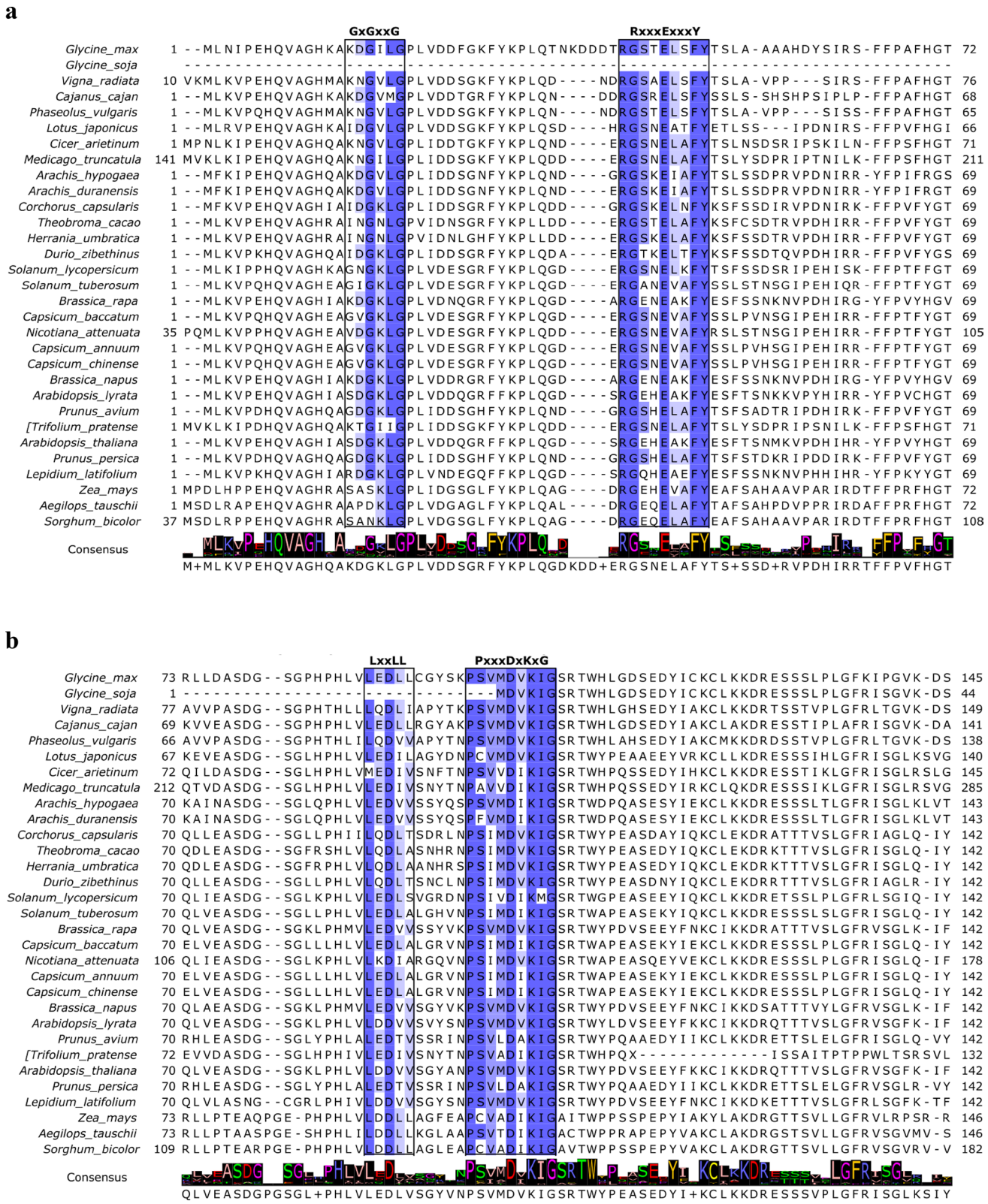


Fig. 4 M-Coffee *multiple* sequence alignment diagram of selected plant IPK2 protein sequences rendered with Jalview. The sequence motifs shared amongst all the representatives are coloured according to percentage identity. The consensus row at the bottom shows the most frequent residue at each column or a '+' if two or more residues are equally abundant

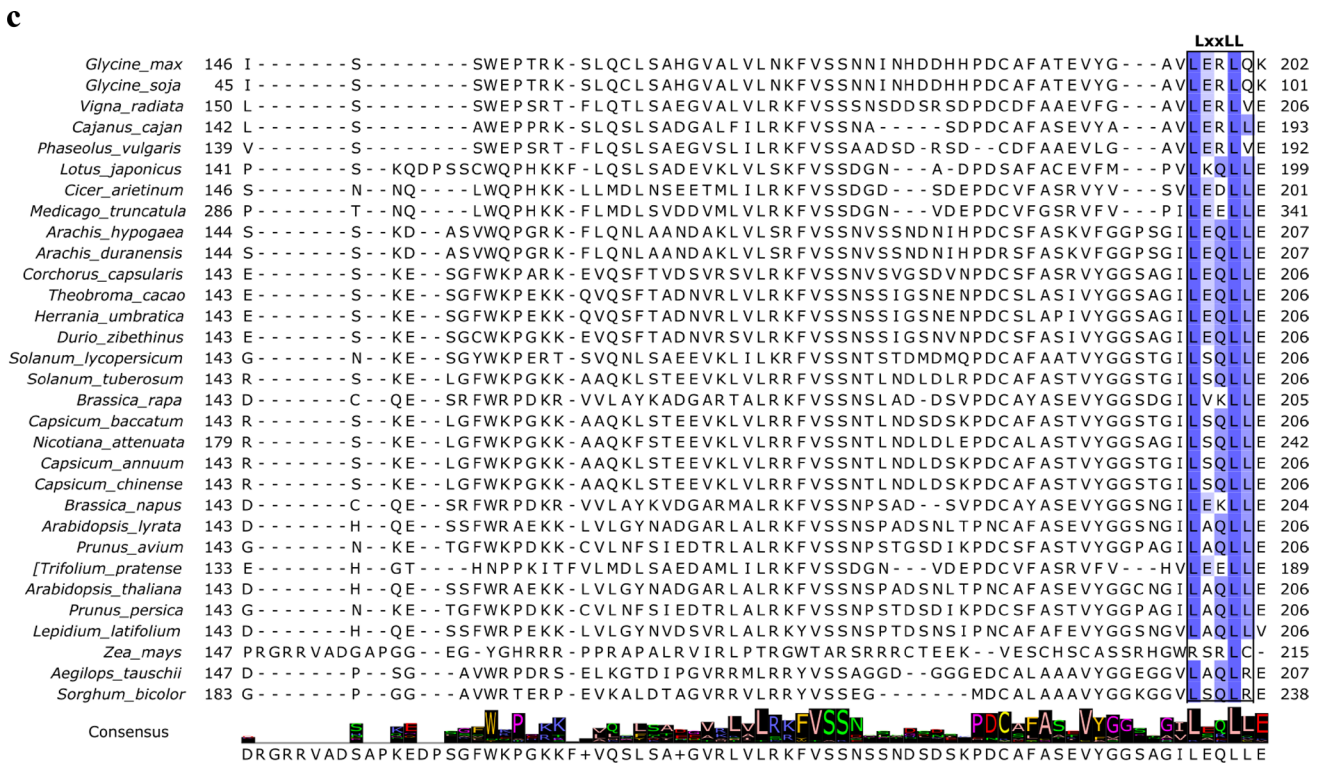


Fig. 4 (continued)

characterization" of GmIPK2 protein and its homologs, we constructed a neighbor-joining phylogenetic tree using MEGA 6.0 software (Fig. 6). The tree topology derived was supported by high bootstrap values. The IPK2s were clustered into six well delineated groups. The clusters consist of members of the Poaceae, Brassicaceae, Malvaceae, Rosaceae, Solanaceae and Fabaceae families. The Poaceae family of monocots (*Zea may, Sorghum bicolor and Aegilopus tauschii*) was found to be most distantly related to *G. max* in comparison to the fabaceae family of eudicots (*Glycine soja, Phaseolus vulgaris, Vigna radiata, Cajanus cajan, Trifolium pratense, Cicer arietinum, Medicago truncatula, Lotus japonicas, Arachis duranensis*, and *Arachis hypogaea*) which are most closely related.

Three-dimensional model construction

Model building

To derive structural information about GmIPK2 protein we built its theoretical model by following homology modeling approach since no X-ray crystal or NMR structure of it is available. The homology modeling technique takes advantage of structural conservation found in similar proteins that have evolved from a common ancestor. Yeast IPK2 protein was the first member of inositol multikinase family whose crystal structure was determined (Holmes and Jogl 2006), but it was found to show a very low sequence similarity with GmIPK2 protein. As we know, aligning two sequences can be a difficult process if the sequence similarity is low, we therefore, used the above derived GmIPK2 protein sequence as a query in PSI-BLAST to find more similar sequences amongst the PDB database proteins with resolved 3D structures to use as a potential template. The only closest homologous sequence available in PDB was that of chain A of *Arabidopsis thaliana* inositol phosphate multikinase (PDB:4FRF) which showed 55% sequence identity with an e value of 9e-99. The initial comparative models, i.e., GmIPK2-S and GmIPK2-P were built using fully automated SWISS-MODEL and PHYRE2 servers, respectively, which also identified 4FRF_A as the most reliable template using sensitive hidden markov model searches and used the same as the structural input. A global quality estimation score (GMQE) of 0.62 was provided by SWISS-MODEL which indicates a reasonably reliable structure. Homology model was additionally built using MODELLER 9.16 program from the X-ray crystal structure coordinates of the previously identified template structure (4FRF_A). The software generated five different models by optimizing the objective function of spatial restrains in a cartesian space. Three different energy scores viz. molpdf, DOPE and GA341 were computed for each of these models and compared to one

Table 3 Three-state description of secondary structure content and disulfide pattern prediction of IPK2 sequences

Organism	α-Helix (%Hh)	Extended strands (%Ee)	Random coil (%Cc)	Disulphide bridge prediction
Glycine max	26.88	13.26	31.54	94–121, 123–276, 158–186
Glycine soja	34.27	29.78	35.96	20–117, 22–85, 161–175
Vigna radiata	29.72	17.13	24.48	127–190, 263–281
Cajanus cajan	25.84	13.11	29.21	119–177
Phaseolus vulgaris	24.62	21.21	28.79	116–176, 208–249
Lotus japonicus	25.44	14.49	26.86	94–215, 149–187, 261–283
Cicer arietinum	21.96	14.53	28.72	122–185, 217–279, 260–274
Medicago truncatula	18.54	15.02	31.69	98–262, 325–357
Arachis hypogaea	24.58	14.48	33.67	120–188, 223–288, 269–292
Arachis duranensis	24.58	14.48	33.33	223–288, 269–292
Corchorus capsularis	19.52	20.21	25.34	120–187, 222–267
Theobroma cacao	19.74	16.78	39.14	50–120, 187–222, 267–298, 288–296
Herrania umbratica	18.69	14.75	42.95	120–296, 222–298, 267–288
Durio zibethinus	18.98	22.37	26.10	93–120, 149–222, 267–288
Solanum lycopersicum	18.03	18.37	30.27	120–187, 222–267
Solanum tuberosum	17.40	17.65	26.23	120–187, 222–283
Brassica rapa	13.50	21.17	22.63	144–186, 221–264
Capsicum baccatum	17.65	17.38	30.75	120–187, 222–283
Nicotiana attenuata	21.47	15.76	25.82	156–223, 258–303
Capsicum annuum	19.52	16.58	30.48	120–187, 222–267
Capsicum chinense	20.32	16.31	30.21	120–187, 222–267
Brassica napus	18.25	18.25	24.56	120–220, 144–185, 264–281
Arabidopsis lyrata	22.00	17.00	23.33	120–187, 222–272
Prunus avium	17.44	17.44	27.05	120–222, 156–187
Trifolium pratense	21.61	15.75	31.87	173–205, 203–250
Arabidopsis thaliana	22.00	16.67	24.67	120–187, 197–272
Prunus persica	17.02	17.02	29.79	120–222, 156–187
Lepidium latifolium	17.85	17.17	28.62	78–269, 120–187, 222–285
Zea mays	16.25	10.42	44.17	101–203, 192–200
Aegilops tauschii	28.57	7.67	31.36	110–188, 124–267
Sorghum bicolor	30.12	12.42	26.40	13–300, 137–160

The secondary structure data were generated by joint prediction with SOPMA and PHD while disulfide bonding pattern was determined using DiANNA (DiAminoacid Neural Network Application) 1.1 server

another to select the best 3D structure (Table S3). Model 4 (GmIPK2-M4), with the lowest molpdf and DOPE scores of 1734.48291 and − 27846.58203 respectively, was chosen as the principal conformational structure.

Structure validation

We then assessed the accuracy and reliability of all the three predicted models viz. GmIPK2-S, GmIPK2-P and GmIPK2-M4 using various online diagnostic tools (Table 4). RAMPAGE server which evaluates the 3D-structures based on Ramachandran plot calculations showed variable distribution of torsion angles in all the models. Fig S3a shows Ramachandran plot for GmIPK2-S model, showing 92.5% residues in favourable region and 4.9% in allowed region with just 2.7% residues in outlier region of the plot, which reflects its superior backbone geometry. We then utilized ProSA-web computational engine to analyze overall quality of the models based on their z scores and local quality based on their residue energies. GmIPK2-S model showed a better z score of − 6.56 which is displayed in the energy distribution plot derived from a group of experimentally determined protein structures of similar size. The Z score thus observed was very much within the range of scores typically found for native conformations of this group which indicates a good overall quality of the modelled protein (Fig S3b). Moreover, its residue energy was computed to be largely negative which further reflects that the local regions in the protein are modelled well (Fig S3c). The accuracy of 3D-models was further analyzed from energy profiles obtained by

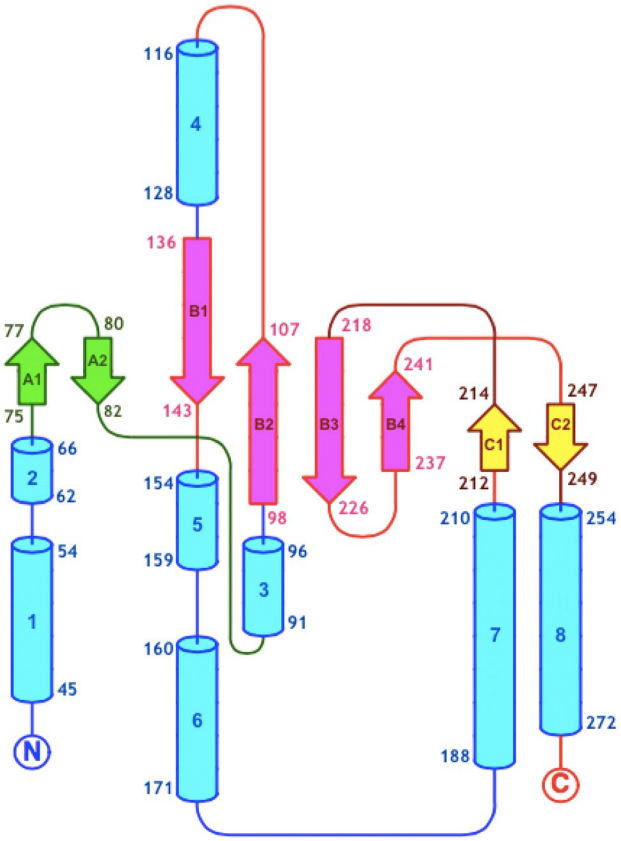


Fig. 5 Topology map of GmIPK2 generated using ProMotif. There are a total of eight α-helices (1–8) and three β-sheets (β-sheet 1: strands A1 and A2; β-sheet 2: strands B1, B2, B3 and B4; β-sheet 3: strands C1 and C2)

Verify3D program. Figure S3(d) shows 3D-1D profile for GmIPK2-S model with 97.82% of the model residues showing an average 3D-1D profile score >= 0.2 and hence validate that majority of its amino acid sequence reconcile to its environment in the 3D structure. Additionally, the pairwise 3D structural alignment of GmIPK2-S model with template protein, 4FRF_A using MATRAS 2.1 program revealed that both the structures shared 91.3% secondary structure identity (Fig S4a) and the average distance between the Cα backbone atoms of their 3D structures, i.e., root mean square deviation (RMSD) measured through superimposition was 0.44 Å (between 196 atom pairs) (Fig S4b). Based on the majority of winning scores, GmIPK2-S (Fig. 7) was chosen as the best comparative model for energy minimization and further analyses.

Molecular dynamics simulation

We subsequently performed MD simulation on our predicted GmIPK2-S model using GROMOS96 53A6 force field to compute its stability and dynamics. Initial potential energy minimization of solvated model showed that the maximum force dropped below the defined value of 1000 kJ mol^{-1} nm^{-1} in nearly 500 steps. The protein structure was then subjected to 50 ns of equilibration run at a constant temperature and pressure to obtain its molecular trajectory. The trajectory thus obtained was used to determine RMSD of Cα backbone atoms of the model using its starting structure as the reference, to determine its convergence towards an equilibrium state. Figure 8a shows RMSD as a function of simulation time. We observed that the protein stabilized around 10 ns of production run and converged to ~0.55 nm at 50 ns. The initial increase of RMSD could be attributed to the restraints in the system applied in the equilibration phase and their release later at the beginning of production phase. Besides RMSD, we also calculated root mean square fluctuations (RMSFs) to study mobility of the protein structure to draw an idea of its flexibility regions. From the RMSF plot (Fig. 8b), we identified Leu (75, 135, 272), Asp (76, 79, 127), Ala (77), Ser (78, 130–132), Gly (80, 136), His (84),

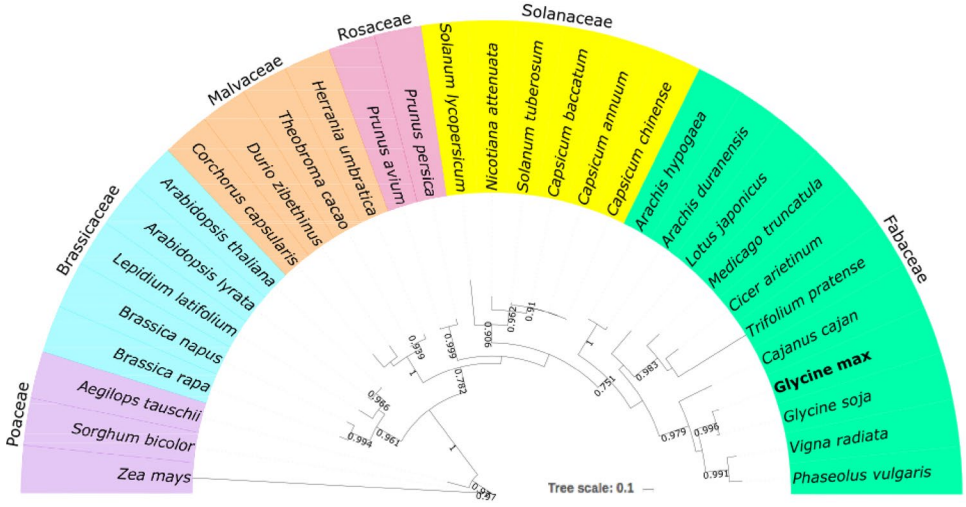


Fig. 6 Phylogenetic tree showing evolutionary relationship of 31 plant IPK2 sequences divided into different clades, colour coded to indicate the plant family to which they belong. The posterior probability values are indicated corresponding to every node

Table 4 Validation parameters computed for the energy-minimized 3D models of GmIPK2 protein built using different programs

Model	RAMPAGE Percentage of residues in favoured region	ProSA *Z* score	Verify3D Percentage of residues with 3D-1D score ≥ 0.2
GmIPK2-S	92.5	−6.56	97.82
GmIPK2-P	82.2	−6.28	84.95
GmIPK2-M4	87.7	−5.84	79.93

Fig. 7 Homology model of GmIPK2_S protein rendered using PyMOL

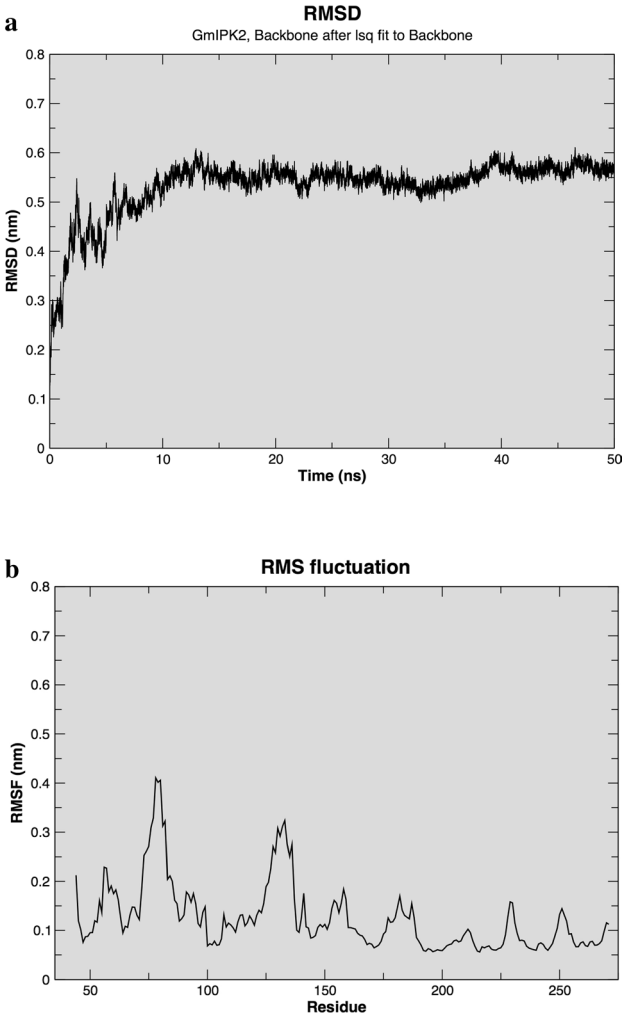


Fig. 8 **a** Root mean square deviations and **b** Root mean square fluctuations of the Cα backbone atoms in GmIPK2_S model over 50 ns MD simulation

Lys (126, 154), Glu (129, 230), Arg (128, 153) residues to be more flexible. This implies that the mentioned residues show greater movement from their native position, i.e., are dynamic in nature and thus are functionally more relevant. Thus, overall, the simulation results highlight the stable and reliable nature of our protein model and find it fit to be used for further active site predictions.

Molecular docking with inositol phosphates

Active site predictions

Once the protein model was refined, a comparative study was performed to detect possible binding pocket residues using CASTp, FTSite and FunFOLD2 binding site prediction servers (Fig S5). Lys105, Thr110, Lys122, Lys126, Ser130, Lys138, Ile139, Pro140, Arg153, Lys154, Gln157, and Ser219 were determined as possible active site residues in GmIPK2-S model. Similar predictions were also made in template protein (4FRF_A) which identified Arg104, Thr105, Pro108, Phe137, Lys149, Arg152, His216, Asn218, Ser219, Gln242, and Val246 as the probable binding site residues. From these studies, we could decipher that residues Lys, Arg, Pro, Thr, Gln and Ser are highly conserved in active sites of functionally identical model and template proteins.

Docking and residue interaction analysis

Molecular recognitions are vital to many biological processes. However, experimental determination of structures of molecular interactions is cost intensive, demand time, and expertise. We therefore, chose computational molecular docking to model our protein–ligand binding and characterize the interactions between its binding pocket residues and

known active ligands. Multiple substrate specificities have been described previously for IPK2 gene product, catalyzing primarily 5GUG-I3P and 4A69-I0P (Stevenson-Paulik et al. 2002). We thus docked these centroid ligands into the binding cavity of GmIPK2 protein using molecular docking program Autodock Vina based on a semi-flexible docking approach with the scaling factor defined within 0.1 nm to predict their bound geometry. VINA uses its iterated local search global optimizer algorithm to produce 9 different poses of which pose 1 corresponding to each ligand was identified as the best binding mode based on their lowest binding affinity score of − 6.2 kcal/mol for 5GUG-I3P and − 5.8 kcal/mol for 4A69-I0P computed by VINA's default statistical scoring function. We further inspected the molecular interactions between these protein substrate poses using Discovery Studio to predict functionally important amino acid residues and found that both the ligands were stabilized in their active site area by strong hydrogen-bonding interactions (Fig. 9a, b). Lys105 and Lys126 were identified as H-donors to the phosphate group oxygen of both 5GUG-I3P and 4A69-I0P while Arg153 a H-acceptor for hydrogen bonds formation suggesting that these binding pocket residues may play a pivotal role in enzymes function and protein structure stability. Based on the previous work conducted by Holmes and Jogl, we hypothesize that side chains of these amino acid residues may assist in inducing conformational changes on inositol phosphate binding, enabling the enzyme to interact with differently phosphorylated inositol polyphosphates in different orientations, thus endorsing its substrate versatility. Besides Ser217 was also found to form stabilizing hydrogen bond with I0P ligand. Gln157 and Pro140 were observed to form an unconventional carbon–oxygen hydrogen bond with 5GUG-I3P and 4A69-I0P respectively indicating their possible contribution to ligand binding affinity and ligand recognition (Klaholz and Moras 2002). Lengths and angles of hydrogen bonds stabilizing the GmIPK2-I3P and GmIPK2-I0P complexes are enlisted in Table 5a, b respectively. Moreover, our analysis indicates that Thr110, Lys138 and Lys154 show non-bonding interactions with the ligands. 2D protein GmIPK2-I3P/I0P ligand

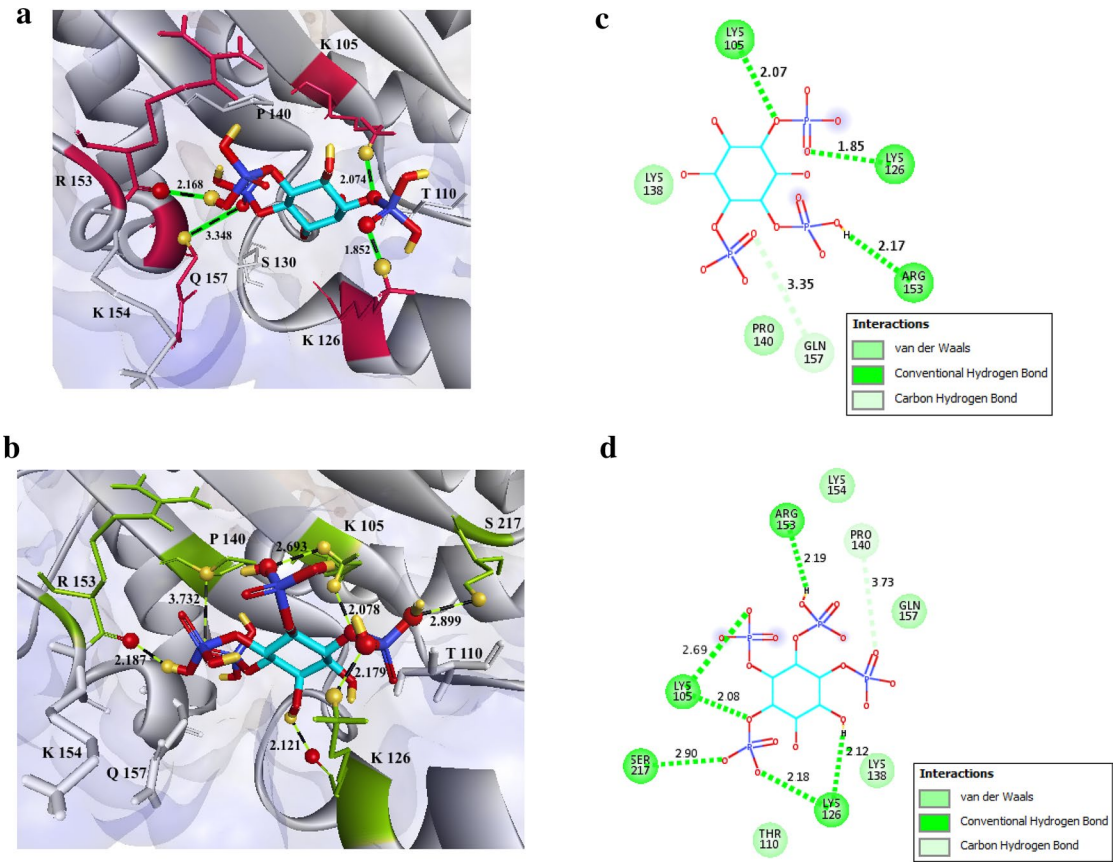


Fig. 9 Molecular docking of substrates to the GmIPK2_S homology model. Hydrogen-bonding interactions of **a** 5GUG-I3P and **b** 4A69-I0P substrates with residues in the active site of GmIPK2_S protein. The substrate is depicted in stick representation with carbon atoms coloured turquoise, oxygen atoms red, hydrogen atoms yellow and phosphorus atoms blue. The interacting protein side chains are represented as maroon sticks in **a** and green sticks in **b. c, d** 2D-schematic representation of the interactions shown in **a, b** respectively drawn using Discovery Studio Visualizer

interaction diagrams (Fig. 9c, d) were also generated using Discovery Studio.

Molecular dynamics simulation of GmIPK2-I3P and GmIPK2-I0P complexes

We subsequently subjected our docked protein–ligand complexes to MD simulation using GROMOS96 43A1 force field to understand their stability and dynamics. The trajectories obtained were utilized to construct their respective RMSDs, RMSFs and H-bond interactions. The GmIPK2-I3P and GmIPK2-I0P complexes exhibited a deviation between ~0.45–0.65 nm (Fig S6) and ~0.45–0.58 nm (Fig S7a), respectively, that converged to ~0.6 and ~0.55 nm, respectively, at 50 ns. This suggests that the structures were stabilized following simulation. From their RMSF plots, we observed fluctuations up to ~0.38 nm in the GmIPK2-I3P complex (Fig S6b) and ~0.29 nm in the GmIPK2-I0P complex (Fig S7(b)) which reveal the characteristic regional flexibilities of functional significance in each complex. We also analyzed the hydrogen bonds which participate in the maintenance of these complexes. Three main hydrogen bonds help to stabilize the ligands, 5GUG-I3P and 4A69-I0P, within the enzyme's active site, two of them acting as acceptors and only one as a donor. In particular, the hydrogen bonds formed with Lys105 and Lys126 and the oxygen atom of each ligand's phosphate group, were constantly held. Figure 10 shows the time evolution of these main H-bonds. Other less significant hydrogen bonds were formed sporadically between the ligands and the protein molecule. Lastly, we calculated the binding free energies ($\Delta G_{binding}$) of both the complexes using *g_mmpbsa* for their respective trajectories. $\Delta G_{binding}$ values of − 12.4 and − 11.9 kcal/mol were estimated for GmIPK2-I3P and GmIPK2-I0P complexes respectively. This indicates that the binding of ligand molecules is thermodynamically favourable and thus validates the reliability of our simulation.

Conclusion

The present work describes expression analysis and molecular characterization of *IPK2* multifunctional kinase involved in the late phase of PA biosynthesis. To initiate the study, we identified and cloned the partial gene sequence of *IPK2* from *G. max* cv. Pusa-16. *IPK2* transcripts, when assessed during seed development showed a predominant expression in the later stages of its development which consequently suggests that perturbing *IPK2* gene at this stage can be a viable strategy for manipulating PA levels in soybean seeds to achieve a *lpa* trait. Computational analysis of the gene further highlighted its molecular features including cis-acting promoter elements potentially regulating its observed expression; primary and secondary structural features shedding light on its physicochemical features, conserved functional motifs and major structural elements; and its evolutionary relationship, which can be used to design several experimental studies.

Table 5 Hydrogen bonds between the active site residues of GmIPK2 and its substrates (a) 5GUG-I3P and (b) 4A69-I0P along with their distances and angles measured using Accelrys Discovery Studio Visualizer 4.1

(a)						
GmIPK2			5GUG-I3P		Distance (A°)	Angle (Degree°)
Residue	Atom	Chemistry	Atom	Chemistry		
Lys105	HZ3	H-Donor	O7	H-Acceptor	2.074	49.9029
Lys126	HZ3	H-Donor	O14	H-Acceptor	1.852	44.2482
Arg153	O	H-Acceptor	H39	H-Donor	2.168	45.6080
Gln157	CA	H-Donor	O18	H-Acceptor	3.348	52.1869
(b)						
GmIPK2			4A69-I0P		Distance (A°)	Angle (Degree°)
Residue	Atom	Chemistry	Atom	Chemistry		
Lys105	HZ2	H-Donor	O18	H-Acceptor	2.693	47.8602
Lys105	HZ3	H-Donor	O24	H-Acceptor	2.078	28.0171
Lys126	HZ3	H-Donor	O28	H-Acceptor	2.179	35.7367
Lys126	O	H-Acceptor	H40	H-Donor	2.121	39.3959
Pro140	CA	H-Donor	O4	H-Acceptor	3.732	60.3277
Arg153	O	H-Acceptor	H36	H-Donor	2.187	16.0359
Ser217	HG	H-Donor	O27	H-Acceptor	2.899	22.5987

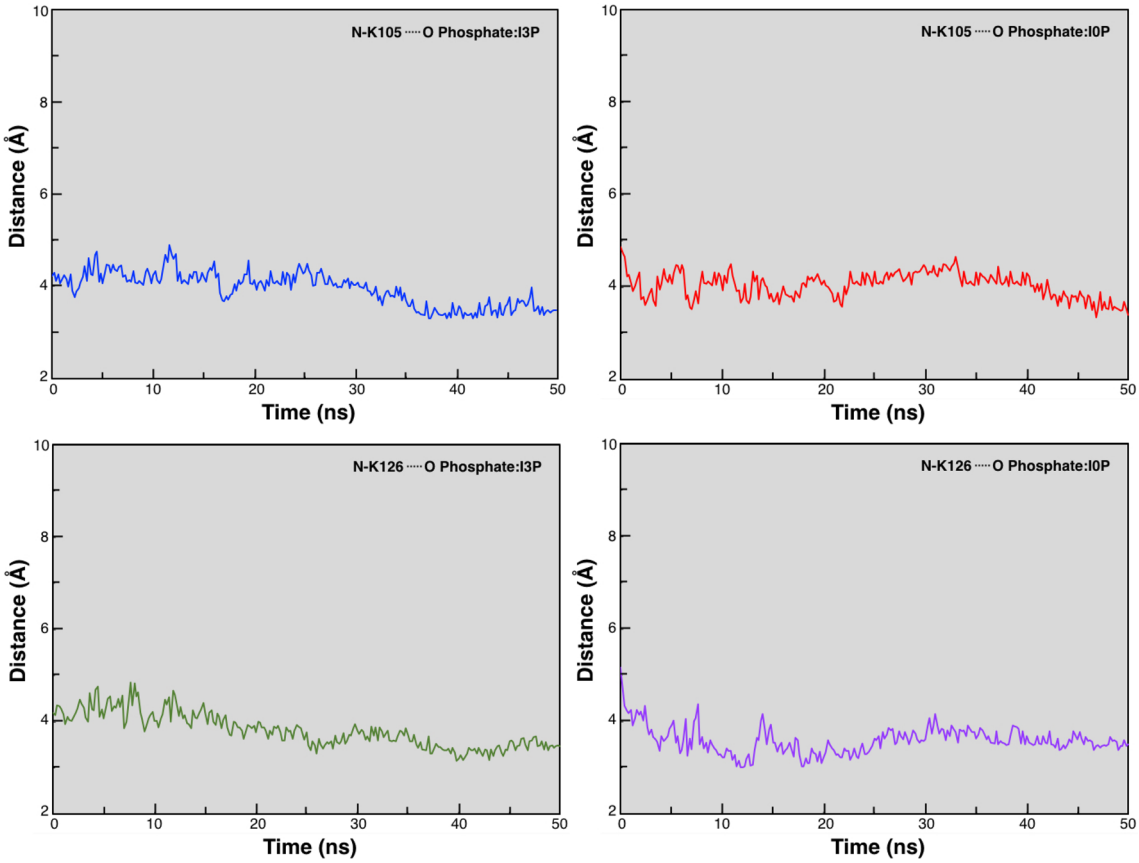


Fig. 10 Time evolution of main hydrogen bonds formed between the GmIPK2_S protein and the 5GUG-I3P and 4A69-I0P ligands over 50 ns MD simulation

The 3D model of IPK2 protein developed can assist in the experimental determination of its 3D model as well as can be considered as a working model for generating hypothesis to make more accurate predictions of protein function and catalytic mechanism in the future. The docking studies performed subsequently can be used for structure-based designing of potent IPK2 inhibitors, useful in studying IPK2 active site and substrate selectivity and for studying PA biosynthesis pathway in detail. In conclusion, the obtained results provide very important preliminary data needed to manipulate PA content in soybean seeds as well as other crops for improving their nutritional quality by biotechnological intervention in the future.

Acknowledgements Financial support for the work was provided by funding from the National Funds for Basic, Strategic and Frontier Application Research in Agriculture [Grant No. NFBSFARA/RNAi-2011/2011-12], ICAR, Government of India. The authors would also like to thankfully acknowledge the Supercomputing Facility for Bioinformatics and Computational Biology at IIT Delhi for the use of its facilities.

Compliance with ethical standards

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

References

Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signalling. Plant Cell 15:63–78

Accelrys Software Inc (2013) Discovery studio modeling environment, release 4.0. Accelrys Software Inc., San Diego

Agarwal R, Mumtaz H, Ali N (2009) Role of inositol polyphosphates in programmed cell death. Mol Cell Biochem 328:155–165

Aggarwal S, Shukla V, Bhati KK, Kaur M, Sharma S, Singh A, Mantri S, Pandey AK (2015) Hormonal regulation and expression profiles of wheat genes involved during phytic acid biosynthesis pathway. Plants (Basel) 4:298–319

Ali N, Paul S, Gayen D, Sarkar SN, Datta SK, Datta K (2013a) 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:e68161

Ali N, Paul S, Gayen D, Sarkar SN, Datta SK, Datta K (2013b) RNAi mediated down regulation of myo-inositol-3-phosphate synthase to generate low phytate rice. PLoS One 6:12

Arguello-Astorga GR, Herrera-Estrella LR (1996) Ancestral multipartite units in light-responsive plant promoters have structural features correlating with specific phototransduction pathways. Plant Physiol 112:1151–1166

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37:W202–W208

Bhati KK, Aggarwal S, Sharma S, Mantri S, Singh S, Bhalla S, Kaur J, Tiwari S, Roy J, 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

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, Schwede T (2014) SWISS-MODEL: modeling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42:W252–W258

Bibikova TN, Zhigiler A, Gilroy S (1997) Root hair growth in Arabidopsis thaliana is directed by calcium and an endogenous polarity. Planta 203:495–505

Bilyeu Kristin D, Zeng P, Coello P, Zhang Zhanyuan J, Krishnan Hari B, Bailey A, Beuselinck Paul R, Polacco Joe C (2008) Quantitative conversion of phytate to inorganic phosphorus in soybean seeds expressing a bacterial phytase. Plant Physiol 146:468–477

Bond CS (2003) TopDraw: a sketchpad for protein structure topology cartoons. Bioinformatics 19:311–312

Brearley CA, Hanke DE (1996a) Inositol phosphates in barley (*Hordeum vulgare* L.) aleurone tissue are stereochemical similar to the products of breakdown of Ins P6 in vitro by wheat bran phytase. Biochem J 318:279–286

Brearley CA, Hanke DE (1996b) Metabolic evidence for the order of addition of individual phosphate esters to the myo-inositol moiety of inositol hexakisphosphate in the duckweed Spirodela polyrhiza L. Biochem J 314:227–233

Buchan DWA, Minneci F, Nugent TCO, Bryson K, Jones DT (2013) Scalable web services for the PSIPRED protein analysis workbench. Nucleic Acids Res 41:W340–W348

Chou KC, Shen HB (2007) MemType-2L: a Web server for predicting membrane proteins and their types by incorporating evolution information through Pse-PSSM. Biochem Biophys Res Commun 360:339–345

Cichy K, Raboy V (2009) Evaluation and development of low-phytate crops. In: Hari Krishnan B (eds) Crop Science Society of America, Soil Science Society of America. American Society of Agronomy, Madison, pp 177–200

Cooper JA, Esch FS, Taylor SS, Hunter T (1984) Phosphorylation sites in enolase and lactate dehydrogenase utilized by tyrosine protein kinases in vivo and in vitro. Biol Chem 259:7835–7841

Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J (2006) CASTp: computed atas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res 34:W116–W118

Eisenberg D, Lüthy R, Bowie JU (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol 277:396–404

Emanuelsson O, Nielsen H, Brunak S, Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. Mol Biol 300:1005–1016

Feldbrugge M, Hahlbrock K, Weisshaar B (1996) The transcriptional regulator CPRF1: expression analysis and gene structure. Mol Gen Genet 251:619–627

Felle HH, Hepler PK (1997) The cytosolic Ca^{2+} concentration gradient of Sinapis alba root hairs as revealed by Ca^{2+} selective microelectrode tests and fura-dextran ratio imaging. Plant Physiol 114:39–45

Feng X, Yoshida KT (2004) Molecular approaches for producing low-phytic-acid grains in rice. Plant Biotechnol 21:183–189

Ferrè F, Clote P (2005) DiANNA: a web server for disulfide connectivity prediction. Nucleic Acids Res 33:W230–W232

Fileppi M, Galasso I, Tagliabue G, Daminati M, Campion B, Doria E, Sparvoli F (2010) Characterisation of structural genes involved in phytic acid biosynthesis in common bean (*Phaseolus vulgaris* L.). Mol Breed 25:453–470

Franklin-Tong VE, Drobak BK, Allan AC, Trewavas AJ (1996) Growth of pollen tubes of Papaver rhoeas is regulated by a slow moving calcium wave propagated by inositol 1,4,5-trisphosphate. Plant Cell 8:1305–1321

Frederick JP, Mattiske D, Wofford JA, Megosh LC, Drake LY, Chiou ST, Hogan BLM, York JD (2005) An essential role for an inositol polyphosphate multikinase, Ipk2, in mouse embryogenesis and second messenger production. Proc Natl Acad Sci 102:8454–8459

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. In: Walker JM (ed) The proteomics protocols handbook. Humana Press, pp 571–607

Geourjon C, Deléage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 11:681–684

Gill SC, Von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem 182:319–326

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:D1178–D1186

Guruprasad K, Reddy BVB, Pandit MW (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng 4:155–161

Hanakahi LA, Bartlet-Jones M, Chappell C, Pappin D, West SC (2000) Binding of inositol phosphate to DNA-PK and stimulation of double-strand break repair. Cell 102:721–729

Hofmann K, Stoffel W (1993) TMbase—a database of membrane spanning proteins segments. Biol Chem 374:166

Holmes W, Jogl G (2006) Crystal structure of inositol phosphate multikinase 2 and implications for substrate specificity. Biol Chem 281:38109–38116

Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K (2007) WoLF PSORT: protein localization predictor. Nucl Acids Res 35:W585–W587

Hua S, Sun Z (2001) Support vector machine approach for protein subcellular localization prediction. Bioinformatics 17:721–728

Ikai AJ (1980) Thermostability and aliphatic index of globular proteins. J Biochem 88:1895–1898

Irvine RF, Schell MJ (2001) Back in the water: the return of the inositol phosphates. Nat Rev Mol Cell Biol 2:327–338

Josefsen L, Bohn L, Sorensen M, Rasmussen S (2007) Characterization of a multifunctional inositol phosphate kinase from rice and barley belonging to the ATP-grasp superfamily. Gene 397:114–125

Kawabata T (2003) MATRAS: a program for protein 3D structure comparison. Nucleic Acids Res 31:3367–3369

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845–858

Kim SR, Kim Y, An G (1993) Identification of methyl jasmonate and salicylic acid response elements from the nopaline synthase (nos) promoter. Plant Physiol 103:97–103

Klaholz B, Moras D (2002) C–H … O hydrogen bonds in the nuclear receptor RARgamma—a potential tool for drug selectivity. Structure 10:1197–1204

Kozakov D, Grove LE, Hall DR, Bohnuud T, Mottarella SE, Luo L, Xia B, Beglov D, Vajda S (2015) The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. Nat Protoc 10:733–755

Krishnan V, Jain P, Tripathi V, 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 transgenics. Plant Omics 8:381–391

Kuwano M, Mimura T, Takaiwa F, Yoshida KT (2009) Generation of stable 'low phytic acid' transgenic rice through antisense repression of the 1D-myo-inositol 3-phosphate synthase gene using the 18-kDa oleosin promoter. Plant Biotechnol 7:96–105

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132

Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res 30:325–327

Letunic I, Bork P (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucl Acids Res 44:W242–W245

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using realtime quantitative PCR and the $2\Delta\Delta C(T)$ method. Methods 25(4):402–408

Lois R, Dietrich A, Hahlbrock K, Schulz W (1989) A phenylalanine ammonia-lyase gene from parsley: structure, regulation and identification of elicitor and light responsive cis-acting elements. EMBO J 8:1641–1648

Lopez-Ochoa L, Acevedo-Hernández G, Martínez-Hernández A, Argüello-Astorga G, Herrera-Estrella L (2007) Structural relationships between diverse cis-acting elements are critical for the functional properties of a rbcS minimal light regulatory unit. J Exp Bot 58:4397–4406

Lovell SC, Davis IW, Arendall WB III, Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by C_α geometry: Φ,Ψ and C_β deviation. Proteins 50:437–450

Majerus PW (1992) Inositol phosphate biochemistry. Annu Rev Biochem 61:225–250

Majerus PW (1996) Inositols do it all. Genes Dev 10:1051–1053

Malho R (1998) Role of 1,4,5-inositol trisphosphate-induced Ca^{2+} release in pollen tube orientation. Sex Plant Reprod 11:231–235

Matsuno K, Fujimura T (2014) Induction of phytic acid synthesis by abscisic acid in suspension-cultured cells of rice. Plant Sci 217–218:152–157

Mongkolsiriwatana C, Pongtongkam P, Peyachoknagul S (2009) In silico promoter analysis of photoperiod-responsive genes identified by DNA microarray in rice (*Oryza sativa* L.). Kasetsart J (Nat Sci) 43:164–177

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 16:2785–2791

Murphy AM, Otto B, Brearley CA, Carr JP, Hanke DE (2008) A role for inositol hexakisphosphate in the maintenance of basal resistance to plant pathogens. Plant J 56:638–652

Neumann GM, Thomas I, Polya GM (1996) Identification of the site on potato carboxypeptidase inhibitor that is phosphorylated by plant calcium-dependent protein kinase. Plant Sci 114:45–51

Nguyen T, Sherratt PJ, Cecil B, Pickett CB (2003) Regulatory mechanisms controlling gene expression mediated by the antioxidant response element. Annu Rev Pharmacol Toxicol 43:233–260

Nishikawa K, Toker A, Johannes FJ, Songyang Z, Cantley LC (1997) Determination of the specific substrate sequence motifs of protein kinase C isozymes. Biol Chem 272:952–960

Niu J, Wang J, Hu H, Chen Y, An J, Cai J, Sun R, Sheng A, Liu X, Lin S (2016) Cross-talk between freezing response and signalling for regulatory transcriptions of MIR475b and its targets by miR475b promoter in Populous suaveolens. Sci Rep 6:20648

Nunes ACS, Vianna GR, Cuneo F, Amaya-Farfan J, de Capdeville G, Rech EL, Aragao FJL (2006) RNAi mediated silencing of the myo-inositol-1-phosphate synthase gene (GmMIPS1) in transgenic soybean inhibited seed development and reduced phytate content. Planta 224:125–132

Odom AR, Stahlberg A, Wente SR, York JD (2000) A role for nuclear inositol 1,4,5-trisphosphate kinase in transcriptional control. Science 287:2026–2029

Omasits U, Ahrens CH, Müller S, Wollscheid B (2014) Protter: interactive protein feature visualization and integration with experimental proteomic data. Bioinformatics 30:884–886

Onodera Y, Suzuki A, Wu CY, Washida H, Takaiwa F (2001) A rice functional transcriptional activator, RISBZ1, responsible for endosperm-specific expression of storage protein genes through GCN4 motif. J Biol Chem 276:14139–14152

Pagni M, Ioannidis V, Cerutti L, Zahn-Zabal M, Jongeneel CV, Hau J, Martin O, Kuznetsov D, Falquet L (2007) MyHits: improvements to an interactive resource for analyzing protein sequences. Nucleic Acids Res 35:W433–W437

Pashou EE, Litou ZI, Liakopoulos TD, Hamodrakas SJ (2004) WaveTM: wavelet-based transmembrane segment prediction. In Silico Biol 4:127–131

Petersen TN, Brunak S, Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–786

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF chimera—a visualization system for exploratory research and analysis. J Comput Chem 25(13):1605–1612

Pierson ES, Miller DD, Callaham DA, Shipley MA, Rivers BA, Cresti M, Hepler PK (1994) Pollen tube growth is coupled to the extracellular calcium ion flux and the intracellular calcium gradient: effect of BAPTA-type buffers and hypertonic media. Plant Cell 6:1815–1828

Plevin MJ, Mills MM, Ikura M (2005) The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. Trends Biochem Sci 30:66–69

Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, Spoel D, Hess B, Lindahl E (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29:845–854

Raboy V, Dickinson DB (1987) The timing and rate of phytic acid accumulation in developing soybean seeds. Plant Physiol 85:841–844

Reehana N, Ahamed AP, Ali DM, Suresh A, Kumar RA, Thajuddin N (2013) Structure based computational analysis and molecular phylogeny of C-Phycocyanin gene from the selected cynobacteria. Int J Biol Vet Agric Food Eng 7:47–51

Rice P, Longden I, Bleasby A (2000) Emboss: the European molecular biology open software suite. Trends Genet 16:276–277

Roberts DM, Harmon AC (1992) Calcium modulated proteins: targets of intracellular calcium signals in higher plants. Annu Rev Plant Physiol Plant Mol Biol 43:375–414

Roche DB, Buenavista MT, McGuffin LJ (2013) The FunFOLD2 server for the prediction of protein-ligand interactions. Nucleic Acids Res 41:W303–W307

Rogers S, Wells R, Rechsteiner M (1986) Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. Science 234:364–368

Rost B, Sander C (1993) Prediction of protein secondary structure at better than 70% accuracy. Mol Biol 232:584–599

Rost B, Casadio R, Fariselli P (1996) Refining neural network predictions for helical transmembrane proteins by dynamic programming. Proc Int Conf Intell Syst Mol Biol 4:192–200

Rouster J, Leah R, Mundy J, Cameron-Mills V (1997) Identification of a methyl jasmonate-responsive region in the promoter of a lipoxygenase 1 gene expressed in barley grain. Plant J 11:513–523

Saiardi A, Erdjument-Bromage H, Snowman A, Tempst P, Snyder SH (1999) Synthesis of diphosphoinositol pentakisphosphate by a newly identified family of higher inositol polyphosphate kinases. Curr Biol 9:1323–1326

Saiardi A, Caffrey JJ, Snyder SH, Shears SB (2000) Inositol polyphosphate multikinase (ArgRIII) determines nuclear mRNA export in *Saccharomyces cerevisiae*. Biol Chem J 275:24686–24692

Schwartz R, Ting CS, King J (2001) Whole proteome pI values correlate with subcellular localizations of proteins for organisms within the three domains of life. Genome Res 11(5):703–709

Sessa G, Meller Y, Fluhr R (1995) A GCC element and a G-box motif participate in ethylene-induced expression of the PRB-1b gene. Plant Mol Biol 28:145–153

Shears SB (2004) How versatile are inositol phosphate kinases? Biochem J 377:265–280

Shen X, Xiao H, Ranallo R, Wu WH, Wu C (2003) Modulation of ATP-dependent chromatin-remodeling complexes by inositol polyphosphates. Science 299:112–114

Shinozaki K, Yamaguchi-Shinozaki K (2000) Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. Curr Opin Plant Biol 3:217–223

Steger DJ, Haswell ES, Miller AL, Wente SR, O'Shea EK (2003) Regulation of chromatin remodeling by inositol polyphosphates. Science 299:114–116

Steinberg SF (2008) Structural basis of protein kinase C isoform function. Physiol Rev 88:1341–1378

Stevenson-Paulik J, Odom AR, York JD (2002) Molecular and biochemical characterization of two plant inositol polyphosphate 6-/3-/5-kinases. Biol Chem J 277:42711–42718

Stevenson-Paulik J, Bastidas GJ, Chiou ST, Frye RA, York JD (2005) Generation of phytate-free seeds in Arabidopsis through disruption of inositol polyphosphate kinases. Proc Natl Acad Sci 102:12612–12617

Stiles AR (2007) Identification and characterization of late pathway enzymes in phytic acid biosynthesis in Glycine max, Dissertation, Virginia Polytechnic Institute and State University

Stiles A, Qian X, Shears S, Grabau E (2008) Metabolic and signaling properties of an ITPK gene family in Glycine max. FEBS Lett 582:1853–1858

Sugimoto T, Kawasaki T, Kato T, Whittier RF, Shibata D, Kawamura Y (1992) cDNA sequence and expression of a phosphoenolpyruvate carboxylase gene from soybean. Plant Mol Biol 20:743–747

Sun Y, Thompson M, Lin G, Butler H, Gao Z, Thornburgh S, Yau K, Smith D, Shukla V (2007) Inositol 1,3,4,5,6-pentakisphosphate 2-kinase from maize: Molecular and biochemical characterization. Plant Physiol 144:1278–1291

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

Sweetman D, Johnson S, Caddick S, Hanke D, Brearley C (2006) Characterization of an Arabidopsis inositol 1,3,4,5,6-pentakisphosphate 2-kinase (AtIPK1). Biochem J 394:95–103

Sweetman D, Stavridou I, Johnson S, Green P, Caddick S, Brearley C (2007) Arabidopsis thaliana inositol 1,3,4-trisphosphate 5/6-kinase 4 (AtITPK4) is an outlier to a family of ATP-grasp fold proteins from Arabidopsis. FEBS Lett 581:4165–4171

Takaiwa F, Yamanouchi U, Yoshihara T, Washida H, Tanabe F, Kato A, Yamada K (1996) Characterization of common cis-regulatory elements responsible for the endosperm-specific expression of members of the rice glutelin multigene family. Plant Mol Biol 30:1207–1221

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. Comput Chem 31:455–461

Turner PJ (2005) XMGRACE, version 5.1.19. center for coastal and land-margin research. Oregon Graduate Institute of Science and Technology, Beaverton

Tuteja JH, Clough SJ, Chan WC, Vodkin LO (2004) Tissue-specific gene silencing mediated by a naturally occurring chalcone synthase gene cluster in glycine max. Plant Cell 16:819–835

Wallace IM, Sullivan O, Higgins DG, Notredame C (2006) M-Coffee: combining multiple sequence alignment methods with T-Coffee. Nucleic Acids Res 34:1692–1699

Washida H, Wu CY, Suzuki A, Yamanouchi U, Akihama T, Harada K, Takaiwa F (1999) Identification of cis-regulatory elements required for endosperm expression of the rice storage protein glutelin gene GluB-1. Plant Mol Biol 40:1–12

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25(9):1189–1191

Webb B, Sali A (2016) Comparative protein structure modeling using modeller. Curr Protoc Bioinform 54:5.6.1–5.6.37

Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35:W407–W410

Wu CY, Adach T, Hatano T, Washida H, Suzuki A, Takaiwa F (1998) Promoters of rice seed storage protein genes direct endosperm-specific gene expression in transgenic rice. Plant Cell Physiol 39:885–889

Wymer CL, Bibikova TN, Gilroy S (1997) Cytoplasmic free calcium distribution during the development of root hairs of *Arabidopsis thaliana*. Plant J 12:427–439

Xu J, Brearley CA, Lin WH, Wang Y, Ye R, Mueller-Roeber B, Xu ZH, Xue HW (2005) A role of Arabidopsis inositol polyphosphate kinase, AtIPK2α, in pollen germination and root growth. Plant Physiol 137:94–103

Yamaguchi-Shinozaki K, Mundy J, Chua NH (1989) Four tightly linked rab genes are differentially expressed in rice. Plant Mol Biol 14:29–39

York JD, Odom AR, Murphy R, Ives EB, Wente SR (1999) A phospholipase C-dependent inositol polyphosphate kinase pathway required for efficient messenger RNA export. Science 285:96–100

Yu CS, Chen YC, Lu CH, Hwang JK (2006) Prediction of protein subcellular localization. Proteins 64:643–651

Zhang ZB, Yang G, Arana F, Chen Z, Li Y, Xia HJ (2007) Arabidopsis inositol polyphosphate 6-/3-Kinase (AtIpk2β) is involved in axillary shoot branching via auxin signaling. Plant Physiol 144:942–951

frontiers in Plant Science

ORIGINAL RESEARCH
published: 14 June 2018
doi: 10.3389/fpls.2018.00804

Development and Evaluation of Low Phytic Acid Soybean by siRNA Triggered Seed Specific Silencing of *Inositol Polyphosphate 6-/3-/5-Kinase* Gene

*Mansi Punjabi[1,2], Navneeta Bharadvaja[1], Monica Jolly[2], Anil Dahuja[2] and Archana Sachdev[2]**

[1] Department of Biotechnology, Delhi Technological University, New Delhi, India, [2] Division of Biochemistry, Indian Agricultural Research Institute, New Delhi, India

Soybean is one of the leading oilseed crop in the world and is showing a remarkable surge in its utilization in formulating animal feeds and supplements. Its dietary consumption, however, is incongruent with its existing industrial demand due to the presence of anti-nutritional factors in sufficiently large amounts. Phytic acid in particular raises concern as it causes a concomitant loss of indigestible complexed minerals and charged proteins in the waste and results in reduced mineral bioavailability in both livestock and humans. Reducing the seed phytate level thus seems indispensable to overcome the nutritional menace associated with soy grain consumption. In order to conceive our objective we designed and expressed a inositol polyphosphate 6-/3-/5-kinase gene-specific RNAi construct in the seeds of Pusa-16 soybean cultivar. We subsequently conducted a genotypic, phenotypic and biochemical analysis of the developed putative transgenic populations and found very low phytic acid levels, moderate accumulation of inorganic phosphate and elevated mineral content in some lines. These low phytic acid lines did not show any reduction in seedling emergence and displayed an overall good agronomic performance.

Keywords: soybean, low phytic acid, inositol polyphosphate 6-/3-/5-kinase, RNAi silencing, seed-specific, *Agrobacterium*-mediated transformation

OPEN ACCESS

Edited by:
James Lloyd,
Stellenbosch University, South Africa

Reviewed by:
Francesca Sparvoli,
Istituto di Biologia e Biotecnologia Agraria (IBBA), Italy
Ajay Kumar Pandey,
National Agri-Food Biotechnology Institute, India

****Correspondence:***
Archana Sachdev
arcs_bio@yahoo.com

Specialty section:
This article was submitted to Plant Biotechnology, a section of the journal Frontiers in Plant Science

Received: *11 September 2017*
Accepted: *25 May 2018*
Published: *14 June 2018*

Citation:
Punjabi M, Bharadvaja N, Jolly M, Dahuja A and Sachdev A (2018) Development and Evaluation of Low Phytic Acid Soybean by siRNA Triggered Seed Specific Silencing of Inositol Polyphosphate 6-/3-/5-Kinase Gene. Front. Plant Sci. 9:804. doi: 10.3389/fpls.2018.00804

INTRODUCTION

For the last 50 years, world population multiplied more rapidly than ever before, and is expected to grow over a third by 2050. Therefore, one of the major challenges agriculture will face in the coming decades is to meet the food demand of the growing population. Scientific community and farmers worldwide are seeking an inexpensive source of energy dense and nutrient rich alternative to serve the purpose. Soybean crop due to its unique nutrient profile provide as a promising option to ensure food security in future. United States Department of Agriculture (USDA) estimated that the global soybean production 2017–2018 will be 348.04 million metric tons. Despite such huge production, the presence of natural anti-nutrients, such as inositol hexakisphosphate (Phytic acid, PA), protease inhibitors, lectins, saponins amongst few others, has limited its consumption (Jiang et al., 2013). PA particularly summon attention as it accounts for over 75% of the total seed

phosphorous (Raboy et al., 1984) as well as chelate many divalent cations to form insoluble salts at the biological pH (Halsted et al., 1974; Jacobsen and Slotfeldt-Ellingsen, 1983). Its consumption, however, is inevitable as it is associated with protein bodies found in the cotyledon of the soybean seeds (Lott, 1984; Lott et al., 1995; Wada and Lott, 1997). Therefore, feeding on high-phytate soy-based diets are often feared to exacerbate mineral and protein malnutrition. In addition, phytate excreted by livestock grazing on soy-rich forage contribute significantly to the environmental phosphorus load (Brinch-Pedersen et al., 2000). These problems have provided us with a strong impetus to develop low phytic acid (*lpa*) soybean grains which could contribute in its biofortification.

The first generation of *lpa* crops were developed using classical breeding. With advancement in technology these efforts were augmented by the aid of forward genetics. Mutations that block the synthesis or accumulation of PA during seed development have been isolated in the past in a variety of crop species (Maize, Raboy et al., 2000; Shi et al., 2005; Wheat, Guttieri et al., 2004; Barley, Larson et al., 1998; Rasmussen and Hatzack, 1998; and Rice, Larson et al., 2000; Liu et al., 2007). *Lpa* mutant lines CX1834 (Wilcox et al., 2000) and LR33 (Hitz et al., 2002) were developed in soybean and made available for public breeding efforts. Unfortunately, the crops were unsuccessful due to their poor agronomic performance which called in for a different approach.

Research in the past has evidenced plant genetic engineering techniques to hold unprecedented potential for crop improvement. The first transgenic strategy that was followed to develop *lpa* soybean was the accumulation of microbial phytases in their seeds (Li et al., 1997; Denbow et al., 1998). However, the cost and the labor incurred for its processing prior to consumption made it an economically unviable alternative. Successively, a sustainable approach directing expression of soybean phytase during embryo development at the site of PA synthesis/storage was followed (Hegeman and Grabau, 2001; Chiera et al., 2004). But only a maximum reduction of 25% in PA level was attained by this method which is significantly less than what has been achieved previously. Therefore, to further reduce the PA level albeit optimally, we thought of genetically engineering soybean by exploring RNA interference (RNAi) technique of reverse genetics, which has recently been used to generate *lpa* rice (Ali et al., 2013a,b) and wheat (Bhati et al., 2016). Nunes et al., 2006 achieved upto 95% reduction in soybean seed PA by RNAi but at the cost of embryo abortion in the progeny zygotes. The observed effect can partly be accounted for by firstly, the choice of gene and secondly, the promoter used for expression of the transgene. Nunes in order to manipulate PA biosynthesis, followed inhibition of the first step which is suggested to be the most effective strategy. However, suppressing *myo*-inositol-1-phosphate synthase (*GmMIPS*) expression may lead to critical alterations in seed PA biosynthesis (Kuwano et al., 2009), subsequently disturbing the cellular phosphorus and inositol homeostasis. Therefore, in our study we decided to chose one of the late PA pathway enzyme, inositol polyphosphate 6-/3-/5-kinase (*GmIPK2*) as the target gene. Due to multiple specificity of this enzyme, participating in sequential phosphorylation of 1D-myo-inositol-1,4,5-trisphosphate to 1D-myo-inositol-1,3,4,5,6-penta*kis*-phosphate and its strategic position in the PA biosynthesis pathway we hypothesize that it shall be successful in achieving a greater but optimal level of PA reduction. Further, Nunes et al. (2006) used a constitutive promoter CaMV35S which resulted in a strong expression in vegetative tissues other than developing seeds. From this we conclude that the promoter we use for expression should be active only in developing seeds, the storage site of PA. Also, we kept in mind that the promoter we use should have the same temporal and spatial activity in the seed as *GmIPK2* to achieve a critical level of suppression. With regard to this purpose, we selected the promoter of reserve protein vicilin, located in protein bodies which is the same site as *GmIPK2* gene expression to drive our RNAi construct.

So far there is no report on successful recovery of fertile *lpa* transgenics of Indian soybean cultivars. In this study, we were able to generate *lpa* lines of *Glycine max* Pusa 16 that along with reduction in phytate levels displayed an increase in available phosphorous and few other important mineral elements assayed. In addition, these *lpa* transgenics that we developed can provide a valuable system to study seed PA synthesis and can aid in developing markers for the *lpa* trait in future.

MATERIALS AND METHODS

Plant Material

Glycine max [L.] Merr. cv. Pusa-16 procured from Division of Genetics, IARI, New Delhi, India was used for isolation of *GmIPK2* gene and as the recipient genotype for subsequent genetic modification. Quantitative characteristics of this cultivar grown primarily in northern plain and hill zones of India have been described in studies made by Karnwal and Singh (2009) and Ramteke et al. (2010; Ramteke and Murlidharan, 2012).

Construction of RNAi Expression Vector

Based on the transcript sequence of soybean *IPK2* gene available in the plant comparative genomics portal Phytozome v9.1 (Glyma.12G240900) we designed primers to amplify and clone a 305 bp fragment including 250 bp of its 3′ end coding sequence and a conserved sequence of 55 bp from its 3′ untranslated region. This nucleotide fragment was used to amplify and directionally clone sense (*GmIPK2_S;* FP 5′-CG CGGATCCGCGTTGCAGAAGCTCAAG-3′ and RP 5′-TCCCC GCGGGG AGCGACACTAATTCAAG-3′) and anti-sense (*GmIPK2_As;* FP 5′-CCGCTCGAGCGGAACGTC TTCGAGT TC-3′ and RP 5′-CCATCGATGGCGCTGTGATTAAGTTCGT A-3′) strands around a 395 bp spacer of soybean fatty acid oley1Δ12 desaturase gene intron (*GmFad2-1*; FP 5′-TCCCCG CGGGGAAGGTCTGTCTTATTTTGAATC-3′ and RP 5′-CCATCGATGGTATACCGCACTAGT AAACCAC-3′) cloned in pCR2.1-TOPO cloning vector to form a hairpin (ihp) structure which was confirmed by automated sequencing. The binary vector pCWAK that will carry this silencing cassette was constructed inhouse in a previous work by ligating fragments from the expression vector pCW66 and binary vector pAKVS both kindly obtained from Dr. Craig Atkins

Laboratory, University of Western Australia, Australia. The final RNAi construct pCWAK-ipk2 was generated by restriction cloning (*BamHI/XhoI*) of ihp cassette in pCWAK under the control of soybean seed specific *vicilin* gene promoter and terminator (**Figure 1**). The plasmid pCWAK-ipk2 also bears phosphinothricin acetyltransferase (PAT) encoding marker gene *bar* within the T-borders driven by the cauliflower mosaic virus (CaMV) 35S promoter to confer tolerance to the herbicide glufosinate for transgenic selection. After extensive verification of pCWAK-ipk2 by restriction digestion, we subsequently mobilized the binary construct into disarmed *Agrobacterium tumefaciens* strain EHA105 by triparental mating using *Escherichia coli DH5α* (pRK2013) helper strain for further use in preliminary transformation experiments.

Generation of Transgenic Plants Expressing RNAi Construct

Soybean transformation was performed following *Agrobacterium*-mediated cotyledonary-node method described by Zhang et al., 1999 and Paz et al., 2004 with minor modifications. Briefly, sterilized soybean seeds germinated on half B5 medium were excised to derive two cotyledonary explants per seed and infected with *A. tumefaciens* harboring the RNAi construct (**Figure 2A**). Following 3 days of co-cultivation in dark in the presence of thiol compounds (3.3 mM L-cysteine, 1 mM dithiothreitol, 1 mM sodium thiosulfate) to improve cell transformation efficiency (Olhoft and Somers, 2001; Olhoft et al., 2001), the explants were sub-cultured twice for 12–14 days on full B5 medium (shoot induction medium, SIM) supplemented with 4 mg/l glufosinate for selective shoot induction (**Figures 2B,C**). Four weeks later, healthy explants were transferred to full MS medium (shoot elongation medium, SEM) containing 5 mg/l glufosinate for continued selection and sub-cultured in fresh medium every 2 weeks until shootlets approximately 2–3 cm high were developed (**Figures 2D,E**). Elongated shootlets were subsequently induced to develop roots by the application of 2mg/l Indole-3-butyric acid (IBA; auxin phytohormone) supplemented in half MS medium (rooting medium, RM) (**Figure 2F**). Well rooted plantlets were transplanted to soil pots for hardening and grown to maturity under 16 h/8 h light/dark regime at National Phytotron Facility, IARI, Delhi, India (**Figures 2G,H**).

Transgene Integration Analysis

PCR Examination

The putative transformants (T_0) that survived hardening were selected for primary screening of transgene integration through PCR analysis. Total genomic DNA was isolated from their leaves and the leaves of non-transformed control plants using genomic DNA mini kit (plant) following manufacturer's (Geneaid, Taiwan) protocol. *GmIPK2_S* forward (FP 5′-CGCGGATCCGCGTTGC AGAA GCTCAAG-3′) and *GmFAD2-1 reverse* (RP 5′-CCATC GATGGTATACCGCACTAGTAA ACCAC-3′) primers were used to amplify a 700 bp fragment using thermal cycling conditions of 1 min at 94°C; 30 cycles of 30 s at 94°C, 30 s at 62°C, 30 s at 72°C; and a final extension of 10 min at 72°C. The amplicon generated was purified and sequenced for confirmation.

Segregation Analysis

Segregation analysis of the transgene was performed by studying *bar* gene expression in the T_1 progenies of each independent transformation events characterized. Seeds from these events were germinated in pots filled with Vermiculite, Cocopeat and Sand mixture (1:1:1). Leaf tissues were collected from 30 days old seedlings for genomic DNA isolation and PCR amplification was carried out using *bar* specific primers (FP 5′-GAACGACGCCCGGCCGACAT-3′ and RP 5′-GTCCAGC TGCCAGAAACCCAC-3′) under thermal cycling conditions of 1 min at 94°C; 30 cycles of 30 s at 94°C, 30 s at 65°C, 1 min at 72°C; and a final extension of 3 min at 72°C, to amplify a 500 bp fragment.

Analysis by Southern Blotting

Southern hybridization analysis was carried out on plants cultivated from seeds of PCR positive T_0 lines to confirm stable T-DNA integration and to estimate the number of copies inserted. Twenty micrograms of genomic DNA from each sample was digested with *PstI* which does not cut the T-DNA region. *XhoI* digested pCWAK-IPK2 plasmid DNA was used as the positive control. The restricted samples were separated by electrophoresis on 0.8% agarose gel and transferred onto positively charged nylon membrane (Axiva, India) by capillary blotting following the method described by Sambrook et al. (1989). The membrane was UV cross-linked and subsequently hybridized with *bar*-specific probe at 42°C for 16 h, biotin labeled using Biotin DecaLabel DNA Labeling Kit (Thermo Scientific, United States). The hybridized membrane was washed and detected using Biotin Chromogenic Detection Kit according to the protocol described by the manufacturers (Thermo Scientific, United States).

Transcript Analysis by Quantitative Real-Time PCR

Southern positive T_2 plants were further analyzed by semiquantitative RT-PCR and quantitative real-time PCR (qRT-PCR) to estimate *IPK2* transcript levels. Total RNA was isolated from fresh soybean seed (8–10 mm) tissue of T_2 transgenic and non-transgenic plants using TRI Reagent (Sigma-Aldrich, United States). First-strand cDNA was synthesized from 5 μg of DNaseI-treated total RNA using M-MLV Reverse Transcriptase (Promega, United States) following manufacturer's instructions. qRT-PCR analysis was subsequently performed following SYBR Green (DyNAmo Flash SYBER Green qPCR Kit, Thermo Scientific, United States) chemistry on PikoReal 96 Real-Time PCR platform (Thermo Scientific, United States). Gene-specific primers were designed to amplify a 124 bp fragment from its conserved region (*qIPK1*F 5′-GGAGCGCTTGCAGAAGC-3′, *qIPK1*R5′-GACCAGAGGGTTGGT AGC-3′). To normalize the variance among samples, house-keeping gene phosphoenol-pyruvate carboxylase (*PEPCo*) (q*PEPCo*F5′-CATGCACCAAA

FIGURE 1 | Map showing the T-DNA region of binary vector pCWAK-ipk2. RB: right border; V-P: vicilin promoter; IPK2_S: *IPK2* sense fragment; FAD2_1: fatty acid oley1 Δ12 desaturase gene intron; IPK2_AS: *IPK2* antisense fragment; V-T: vicilin terminator; 35S-P: cauliflower mosaic virus 35S promoter; BAR: bialaphos resistance gene; N-T: nopaline synthase terminator; LB: left border.

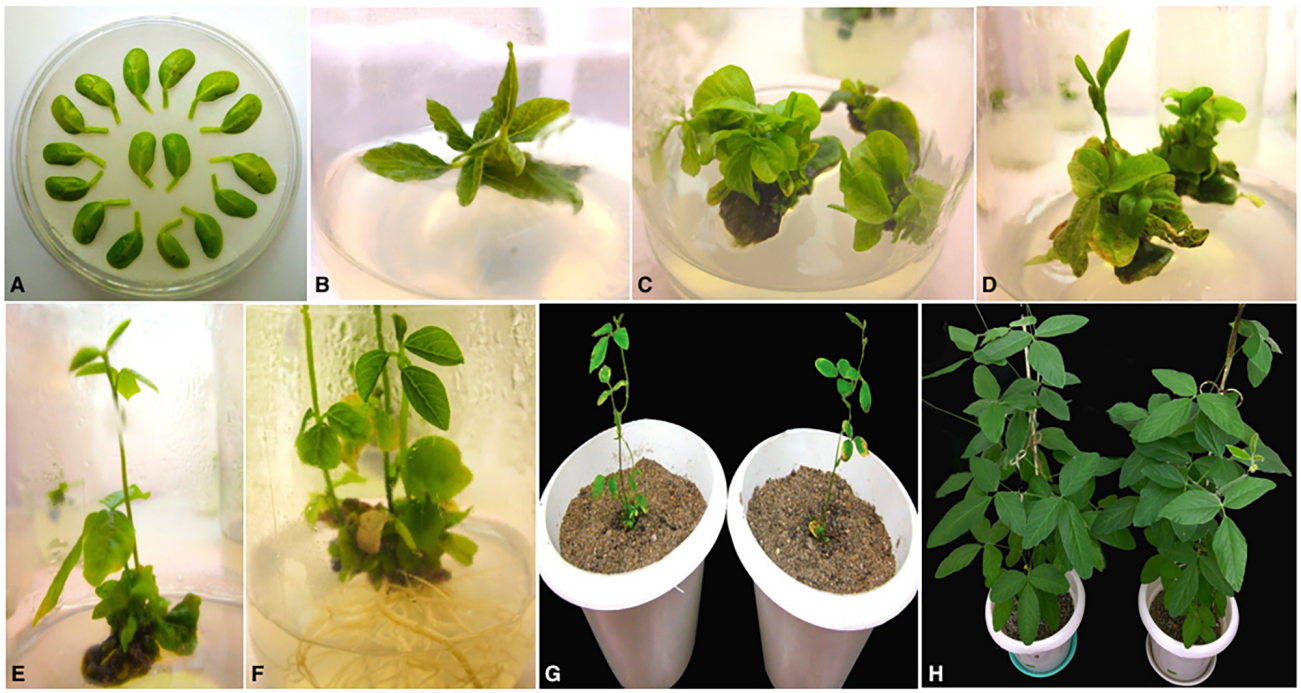


FIGURE 2 | Different stages of *Agrobacterium*-mediated cotyledonary-node transformation of soybean cv. Pusa 16. **(A)** Co-cultured wounded cotyledonary-node explants **(B)** shoot induction from axillary meristem on SIM after 2 weeks; **(C)** shoot development with 4 mg/l glufosinate selection on SIM after 4 weeks of culture; **(D)** shoot elongation with 5 mg/l glufosinate selection on SEM after 6 weeks of culture; **(E)** elongated, glufosinate-resistant shoot ready for transfer to RM after 9 weeks of culture; **(F)** root development on RM; **(G)** acclimatization of plantlets in sterilized pot mix; **(H)** mature T_0 transgenic plants.

GGGTGTTTT-3′, q*PEPCo*R 5′-TTTTGCG GCAGCTATCTC TC-3′) was used as endogenous control. Amplification was achieved by a two-step PCR reaction with an initial denaturation at 94°C for 4 min followed by 40 cycles of 94°C for 30 s and annealing/extension at 60°C for 30 s. Triplicate quantitative assays were performed on biological triplate corresponding to each sample. Relative mRNA abundance was calculated using the $2^{-\Delta\Delta CT}$ method described by Livak and Schmittgen, 2001. The specificity of each unique amplification product was determined by melting curve analysis.

Estimation of Seed Phosphorus Levels

Total phosphorous in the seeds was determined colorimetrically by a method described by Chen et al. (1956). Samples (transgenic and non-transgenic control) for the assay were prepared by wet ashing dried ground seeds (250 mg) in 2 ml of concentrated sulfuric acid (Larson et al., 2000; Raboy et al., 2000).

For the analysis of inorganic phosphorus (Pi) levels, we followed a protocol described by Al-Amery et al. (2015) which uses single seed chips (1–2 mg) derived from cotyledonary segment opposite to the embryonic axis. Briefly, seed samples were extracted using 50 µl buffer (25 mM magnesium chloride and 12.5% trichloroacetic acid) for 14–16 h at 37°C with gentle shaking. 10 µl subsample of extracts were loaded in triplicates onto a 96 well plate and diluted with 90 µl water. Each sample was incubated with 100 µl of Chen's reagent at 37°C for 1 h and the absorbance was read at 882 nm on GloMax®-Multi Detection System (Promega, United States).

Analysis of Phytic Acid Concentration by HPLC

Phytic acid estimation in transgenic seeds was performed by reversed-phase high-performance liquid chromatography (RP-HPLC) as described by Lehrfeld (1989) with some modifications

(Pandey et al., 2016). 500 mg dried seed tissue was homogenized in a pestle-mortar, hexane-defatted and subsequently extracted with 0.78 M HCl by sonication (3 min) and mechanical agitation for 1 h at 250 rpm. The extract was centrifuged and one part of clear supernatant was mixed with four parts of HPLC grade water. The diluted sample was passed through conditioned SAX column (Hypersep, Thermo Scientific, United States) connected to a vacuum manifold (Millipore, United States) and eluted with 2 M HCl. The filtrate was evaporated in a vacuum rotary evaporator (Hei-VAP Value Digital G3, Heidolph, Germany), re-dissolved in 1 ml of mobile phase (acetonitrile, formic acid and tetrabutylammonium hydroxide, 4.8:5.1:0.1, v/v/v) and filtered using a 0.2 micron PVDF syringe filter (Millipore, United States). For HPLC analysis, sample was injected onto the C18 RP-column (250 × 4.6 mm, 5 μ; Shimadzu, Japan) equilibrated with isocratic mobile phase at a flow rate of 1.0 ml min^{-1}. PA signals were monitored with a UV-VIS photodiode array detector (SPDM-20Avp, Shimadzu, Japan) at a wavelength of 197 nm. The PA concentration was calculated using the calibration curve prepared with a PA dodecasodium salt standard (Sigma-Aldrich). The software used for analysis was LC-Solutions (version 1.25).

In Vitro Bioavailability Assay

Mature transgenic and non-transgenic seeds were milled (Wiley, Thomas Scientific, United States), weighed (1 g) and digested in triplicates under simulated gastrointestinal conditions following method described by Kiers et al., 2000 to determine the bioavailability of iron (Fe), zinc (Zn), and calcium (Ca). The digest was centrifuged at 9000 rpm for 15 min and supernatant obtained was passed through a 0.45 mm filter. The filtrate was analyzed for Fe, Zn, and Ca using an Atomic Absorption Spectrophotometer (AAnalyst 200, Perkin Elmer, United States). A blank consisting of distilled water was processed in a similar manner and used for sample correction.

Agronomic Evaluation of Transgenic Plants

Germination Assay for Seed Viability

The germination capacity of T_3 transgenic seeds as compared to non-transgenic control was assessed by performing controlled germination test (CGT) described by Campion et al., 2009. The seeds were immersed under water for 8 h at 28°C and then transferred to fresh water for an additional 12 h. Following incubation, seeds were rinsed three to four times in distilled water and subsequently germinated in petri dishes lined with filter paper soaked in distilled water under a photoperiod of 28°C 8-h dark and 30°C 16-h light. We also monitored seed vigor by using the standard accelerated aging test (AAT) test. In this we subjected unimbibed seeds to conditions of high temperature (42°C) and high relative humidity (90%) for 72 h and then removed them from the stress conditions and placed under optimum germination conditions.

Phenotypic Analysis

Different morphological traits of transgenic plants were evaluated under controlled conditions to capitulate the influence of transgene integration on phenotype. Mature transgenic plants were harvested and evaluated with respect to the non-transgenic control for plant height, number of pods, number of seeds, seed dry weight, stem length and root length. The height of individual plant was measured as the distance from the soil surface to the tip of the plant. On maturity, upto 100 seeds were harvested and their mean dry weight (SDW) was recorded after air-drying them in oven at 60°C for 72–96 h. Five randomly chosen plants from each transgenic line were evaluated for each parameter studied.

Statistical Analysis

Every experiment was carried out in biological and experimental replications (three to six) for each non-transgenic control and transgenic sample and represented as mean ± standard error. All statistical evaluations were performed using SAS software (version 9.2). Segregation patterns analyzed with chi-square (χ^2) goodness of fit were tested at 5% level of significance. We also conducted ANOVA for agronomic traits of transgenics and non-transgenic control and compared the group means by Tukey's test at 5% level of significance.

RESULTS

Transgene Integration and Segregation Analysis

A preliminary screening was conducted on the T_0 plants to identify putative *lpa* transgenic lines by PCR amplification of ∼700 bp fragment of ihp construct. The assay characterized eight independent transformation events each showing the expected amplicon size which was absent in the case of untransformed control plants (**Figure 3**).

Seeds from each independent event identified were grown successfully under containment conditions to obtain T_1 progeny plants by self pollination. All the derived progenies appeared phenotypically normal and fertile. We performed segregation analysis on them by studying *bar* gene expression and found that five of the events (P2, P4, P5, P6, and P8) held a good fit to the stable Mendelian inheritance of a single locus (3:1) (**Table 1**) whilst three (P1, P3, and P7) of them showed a segregation ratio of two transgene loci (15:1).

We further confirmed integration of the transgene cassette by southern blot analysis using *bar* gene probe (**Figure 4**). The genomic DNA was digested with *PstI* endonuclease which is not present within the T-DNA of the construct. Upon detection, all the transgenic lines analyzed showed a distinct pattern of seperation which indicate that each plant originated from an independent transformation event. The blot showed that all the fragments were above 3 kb in size which suggest that the transgenic lines carry intact copies of T-DNA because the shortest fragments in each line were longer than the T-DNA region (∼2.8 kb). The blot also confirmed single copy integration of target gene into the genome of transgenic events P2, P4, P5, P6, and P8. Since single copy insertions are always desirable, T_2 generation plants were cultivated from seeds of these five characterized events and PCR screened to select homozygous lines. A positive amplification in all the tested T_2 plants derived from the T_1 parents, P2-45 and P6-39, suggest their homozygous

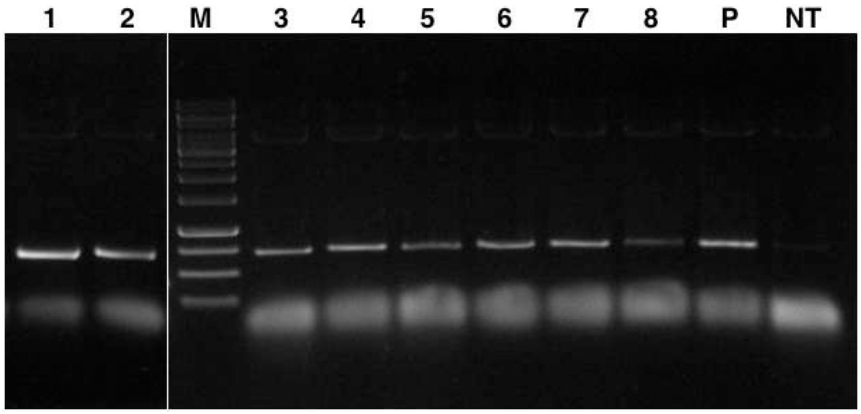


FIGURE 3 | PCR amplification of ∼700 bp *GmIPK2_S* gene plus *GmFAD2-1* intron fragment from genomic DNA of T_0 transgenic plants. Lanes, M: 1 kb DNA ladder; 1–8: genomic DNA from each transformation event characterized; P: pCWAK-ipk2 plasmid DNA (positive control); NT: genomic DNA from non-transformed plant (negative control). Representative lanes were spliced from the original gel (Supplementary Figure S1).

TABLE 1 | Segregation of *bar* gene amongst T_1 progenies of eight independent transformation events characterized in soybean cv. Pusa 16.

Transgenic events	No. of seeds tested (n)	No. of seedlings				Ratio	Chi-square value (χ^2)*
		PCR +		PCR -			
		Obs.	Exp.	Obs.	Exp.		
P1	55	50	51.56	5	3.43	15:1	0.766
P2	47	37	35.25	10	11.75	3:1	0.347
P3	58	53	54.38	5	3.62	15:1	0.561
P4	44	35	33.00	9	11.00	3:1	0.485
P5	49	39	36.75	10	12.25	3:1	0.551
P6	52	36	39.00	16	13.00	3:1	0.923
P7	56	51	52.50	5	3.5	15:1	0.686
P8	53	37	39.75	16	13.25	3:1	0.773

*$\chi^2_{20.05}$; *df1 = 3.841.*

nature and thus seeds collected from only these parents were selected to characterize *lpa* trait by further analysis.

Expression Analysis of Transgenic Plants

To assess the level of reduction in *GmIPK2* expression in transgenic seeds with respect to non-transgenic control, we carried out RT-PCR and qRT-PCR analysis. From RT-PCR results, we observed a variable reduction in the expression of *GmIPK2* gene with no variation in *PEPCo* transcript expression between the non-transgenic and transgenic plants thus supporting its use as a stable endogenous reference (**Figure 5A**). We further quantified the variations in *GmIPK2* transcript levels by qRT-PCR and found a maximum reduction of 2.79-fold and 2.56-fold in T_3 developing seeds from the transgenic lines P2-45-8 and P6-39-10, respectively (**Figure 5B**) at $P \geq 0.05$. We also studied *GmIPK2* expression in other plant tissues of these two lines and found no variation in its transcript levels with respect to the control plants (Supplementary Figures S3, S4) which suggest its successful silencing only in the seeds.

Seed Phytic Acid Quantification

To further examine the effect of *GmIPK2* silencing on PA synthesis, we quantified PA content in mature grain extracts of above two lines and non-transgenic samples by HPLC analysis. The chromatogram obtained by scanning the samples at 197 nm displayed PA retention peaks around 7.006 ± 0.09 min, the area under which was used to compute the level of PA. A decrease in the PA content was observed in T_3 seeds compared to control which displayed larger peak area indicating its higher concentration in the seeds (**Figures 6–8**). The mean PA content was 3.48 g $100g^{-1}$ for non-transgenic seeds against 1.91 g 100 g^{-1} and 2.02 g 100 g^{-1} for the seeds from transgenic lines P2-45-8 and P6-39-10, showing a maximal reduction of 45 and 42%, respectively (**Figure 9**).

Analysis of Seed Phosphorus Levels

We estimated the total and free phosphorus (Pi) levels in transgenic as well as non-transgenic seeds to correlate it with the reduction in PA content. No significant difference in total seed phosphorus (P) was observed for the RNAi lines as compared to

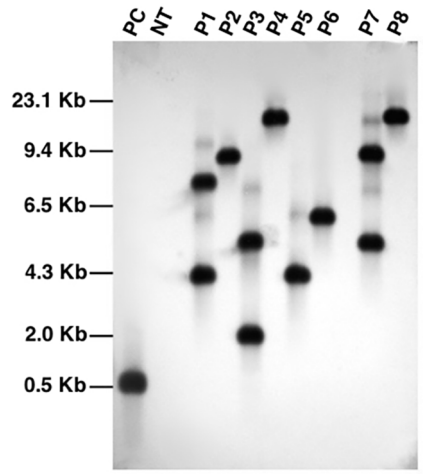


FIGURE 4 | Southern blot analysis of PCR characterized transgenic events using *bar* gene specific probe. Lanes, PC: plasmid pCWAK-ipk2 (positive control); NT: genomic DNA from non-transformed plant (negative control); P1-P8: genomic DNA of independent transgenic events showing hybridization signals indicating transgene integration.

the control. The average total phosphorus content of T_3 seeds was 5.82 mg g^{-1} (P2-45-8) and 5.75 mg g^{-1} (P6-39-10), which was observed closely similar to that of non-transgenic seeds, 5.89 mg g^{-1}. However, a significant increase of 39 and 37% in the concentration of Pi was recorded in the T_3 seeds obtained from transgenic lines P2-45-8 and P6-39-10, respectively, which gave a dark blue reaction with chen's reagent, confirming a reduction in the content of its principal storage form PA (**Figure 9**).

Quantification of Mineral Content in Seeds

In order to determine whether mitigation of PA improved the mineral bioavailability in soy grains, we made a quantitative estimation of micronutrients in transgenic seeds compared to non-transgenic control by following *in vitro* digestion method. From the observations made by atomic absorption spectrophotometer, we noticed an increase in the *in vitro* Fe, Ca, and Zn availability from 54.8, 44.15 and 58.50% for non-transgenic control seeds to 71.1, 59.2, 66.9% for P2-45-8 and 70.7, 58.42, 64.50% for P6-39-10 seeds (**Table 2**). As mentioned before, PA binding with minerals result in formation of insoluble salts with poor bioavailability, our result suggests that the mineral bioavailability can significantly be improved by reduction of PA in *lpa* transgenic lines.

Seed Germination and Morphological Trait Analysis of Transgenic Plants

Decrease in seed PA concentration can be accompanied by adverse agronomic consequences. Therefore, we conducted

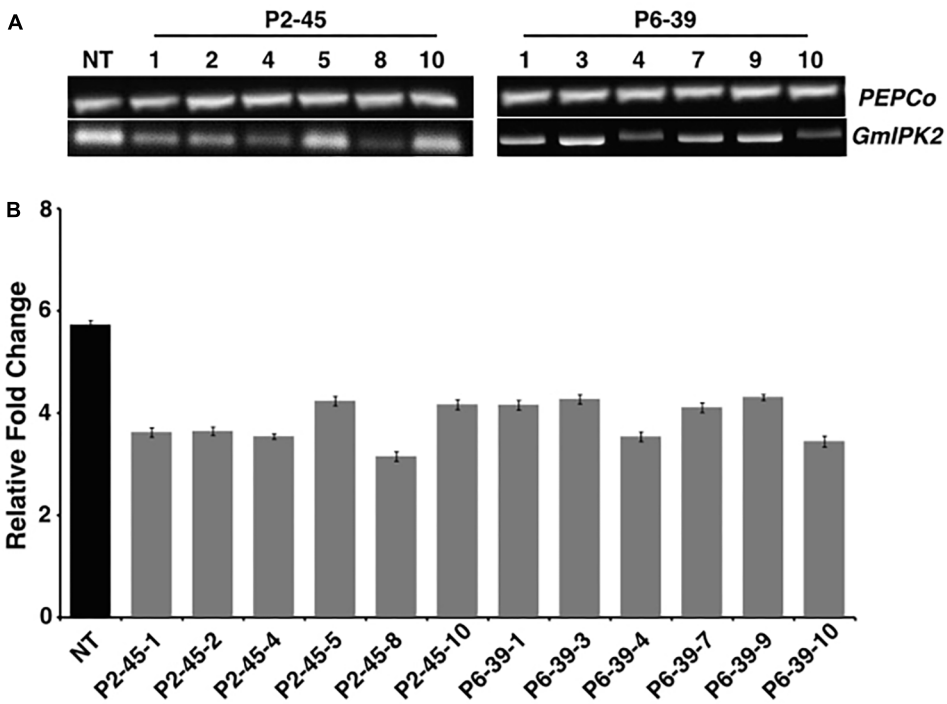


FIGURE 5 | Expression analysis of transgenic soybean seeds. **(A)** RT-PCR amplification showing variation in *GmIPK2* transcripts in T_3 seeds of P2-45 and P6-39 compared to the *PEPCo* internal control (NT: Non-transformed plant). Representative lanes were spliced from the original gel (Supplementary Figure S2). **(B)** Relative fold change measured by qRT-PCR in the above samples indicating varied levels of silencing with maximum reduction observed in P2-45-8. The data presented is mean of technical triplicates corresponding to each biological replicate (n = 3) with error bars indicating standard deviation (SD).

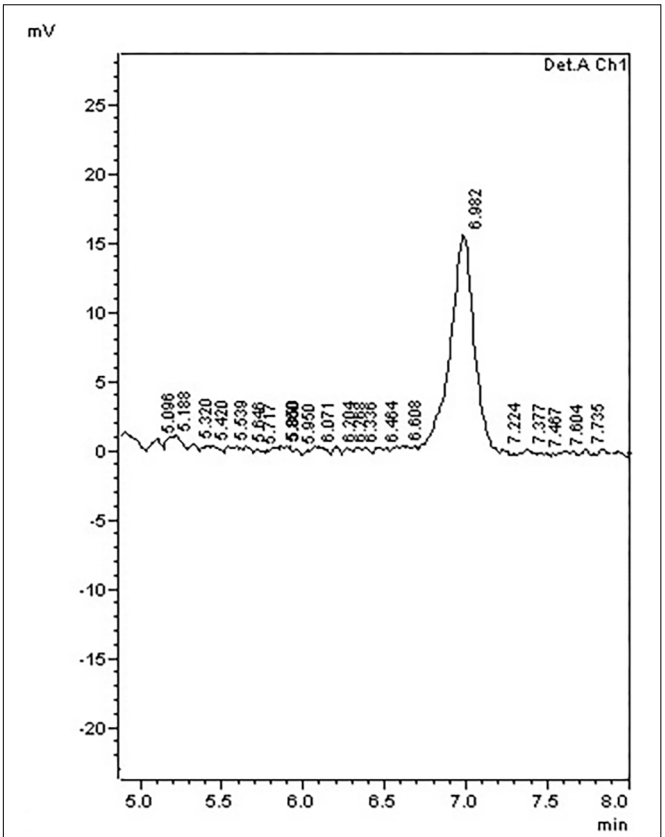


FIGURE 6 | HPLC phytic acid peaks of non-transgenic control seed extracts.

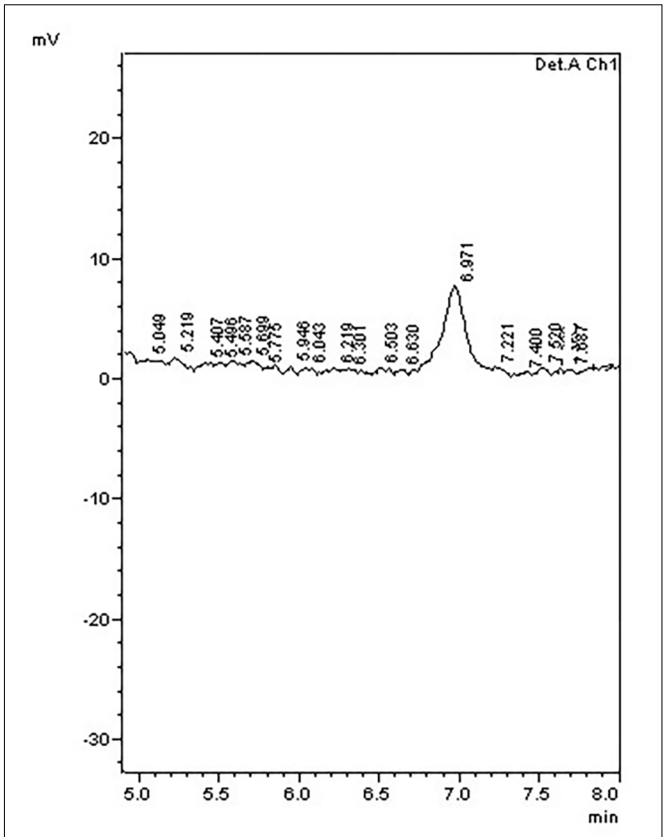


FIGURE 7 | HPLC phytic acid peaks of seed extracts from transgenic line P2-45-8.

germination tests on the *lpa* seeds collected from our transgenic lines. The CGT test so conducted confirmed that 98% of the seeds possessed good germination potential whereas AAT test established their high vigor (Supplementary Figure S5A). Also, the morphological analysis of seed germination in each of the tests revealed similar phenotypes in both transgenic and control seeds (Supplementary Figure S5B). Besides, since PA also regulates the growth and development of seedlings, we carried out phenotyping on 75 days old plants. The growth of transgenic plants was compared with the growth of control soybean plants under field growth conditions. Interestingly, P2 and P6 plants exhibited an increase in leaf size which could possibly be explained due to increase in available seed P content ($P < 0.05$) (Tubana and George, 1997). No significant differences were observed between wild type and transgenic plants in other agronomic traits investigated ($P > 0.05$) (**Table 3**). This data confirms that *GmIPK2* silenced, *lpa* soybean seeds are capable of generating plants showing normal morphologies and agronomic traits.

DISCUSSION

Several mutational and transgenic strategies have been explored to reduce PA in soybean but with undesirable aspects (Meis et al., 2003; Yuan et al., 2007, 2009; Anderson and Fehr, 2008; Maupin and Rainey, 2011; Maupin et al., 2011). We may attribute these negative effects primarily to the constitutive suppression of PA biosynthesis genes which result in an indirect impact on several metabolic processes and signal transduction pathways. Prattley and Stanley (1982) established that the cotyledon of soybean contains 90% of the phytate of the seed, we therefore, expressed our RNAi construct under the control of a cotyledon-specific vicilin promoter (Higgins et al., 1988; Rosche et al., 2002) with the aim to restrict PA perturbations to the seed. As expected, initial characterization of the transgenic genotypes displayed stable integration and inheritance of the transforming DNA with no significant impact on whole plant performance and yield. Segregation patterns observed in the T_1 progenies of distinguished lineages demonstrate that both homozygous and heterozygous plants were obtained from self-pollinated seeds of most of the events, suggesting that T_0 plants were uniformly transformed and were heterozygous for the transformed genotype. Transgenic plants containing single copy or double copies of the transforming gene were also characterized by segregation analysis. Hobbs et al., 1993 through their studies established that single copy transgenic plants are always desirable to avoid the possibility of gene

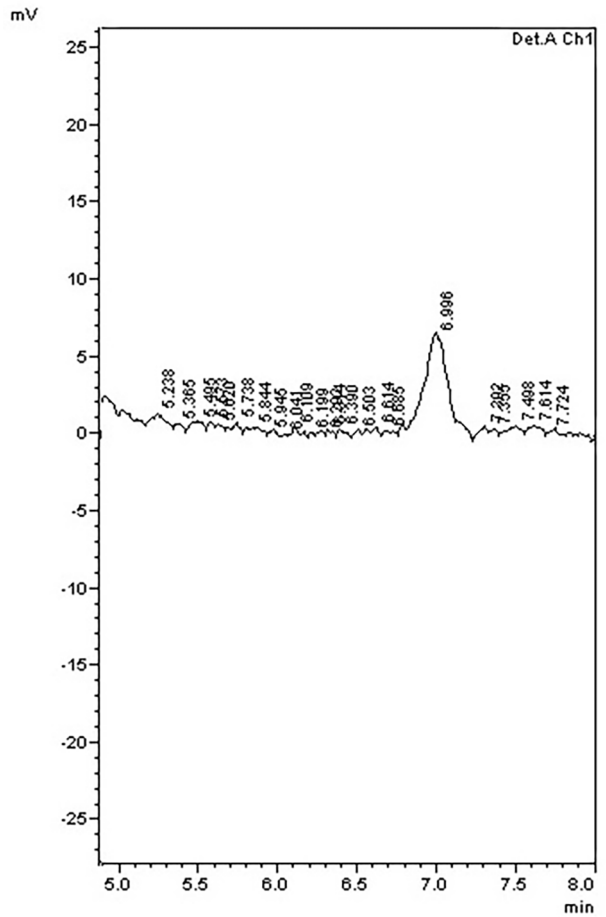


FIGURE 8 | HPLC phytic acid peaks of seed extracts from transgenic line P6-39-10.

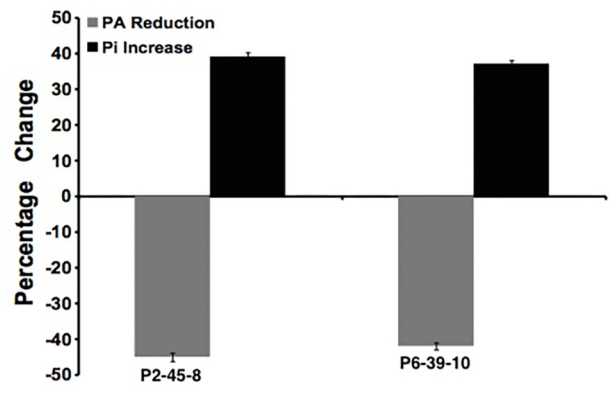


FIGURE 9 | Percentage reduction in phytate (gray column) and increase in free phosphate (black column) in P2-45-8 and P6-39-10 transgenic lines.

silencing during introgression of the transgene into the parental line(s) of hybrids, or into improved open-pollinated varieties, we therefore confirmed these single copy events by performing southern blot before proceeding any further analysis.

TABLE 2 | *In vitro* bioavailability assay of T_3 seeds.

Minerals	*In vitro* availability (%)		
	Non-transgenic (control)	P2-45-8	P6-39-10
Iron	54.80 ± 0.92	71.10 ± 1.11	70.70 ± 1.05
Calcium	44.15 ± 1.04	59.20 ± 0.78	58.42 ± 0.98
Zinc	58.50 ± 0.85	66.90 ± 0.93	64.50 ± 1.07

Data represent means ± SD, n = 3.

TABLE 3 | Morphological and yield contributing characters of T_2 transgenic plants raised in green-house.

Parameters	Non-transgenic (control)	P2-45-8	P6-39-10
Shoot length (cm)	73.40 ± 1.68	72.11 ± 1.42	70.20 ± 1.75
Root length (cm)	30.07 ± 0.15	28.43 ± 0.21	27.64 ± 0.15
Leaf length (cm)	11.20 ± 0.29	12.86 ± 0.18	12.67 ± 0.24
Leaf width (cm)	5.36 ± 0.18	6.01 ± 0.14	5.98 ± 0.17
No. of trifoliates per plant	23.07 ± 0.84	24.07 ± 0.43	23.55 ± 0.26
No. of pods per plant	49.45 ± 0.94	46.50 ± 1.08	45.18 ± 1.24
No. of seeds per pod	2.63 ± 0.08	2.72 ± 0.14	2.54 ± 0.11
100 seeds dry wt (g)	15.21 ± 0.25	14.95 ± 0.14	14.83 ± 0.32

Data represent means ± SD, n = 6.

We used qPCR to check *GmIPK2* expression in T_3 seeds as well as other plant tissues. From our results, we observed a varied level of reduction of *GmIPK2* transcripts in seeds of lines P2-45 and P6-39. These variations can be explained by several factors viz. differences in the site of integration on the chromosomal DNA, possible promoter methylation etc. (Day et al., 2000; Van Leeuwen et al., 2001). On the other hand, no *GmIPK2* reduction was detected in other tissues of transgenic and control plants demonstrating seed-specific regulation. We further confirmed the implications of *GmIPK2* silencing by testing the extracts from T_3 generation seeds for both PA and Pi. A reduction of 45% in the PA content and a 39% increase in Pi levels was found in transgenic line P2-45-8, suggesting a strict correspondence between seed PA and Pi which is in accordance with the previous reports by Larson et al., 2000. Moreover, despite the variations in PA and Pi levels it was found that the mean total P content in the transgenic seeds was unaffected, which suggest that these altered genotypes influence P partitioning to strike a balance for supporting P-related mechanisms in the seed.

Phytic acid represents a major sink for mineral deposition due to its natural strong chelation ability. Research has previously reported the beneficial effects of cereal *lpa* mutants in improved mineral cations bioavailability (Mendoza et al., 1998; Adams et al., 2002; Hambidge et al., 2004, 2005; Linares et al., 2006; Mazariegos et al., 2006; Veum et al., 2007). We therefore analyzed the mineral concentration of transgenic seeds and found that disruption of *GmIPK2* expression caused the Fe, Ca, and Zn contents of seeds to increase. We must also highlight that

amongst the three different minerals analyzed, a maximum increase was observed in the levels of Fe followed by Ca and Zn, respectively, supporting the findings of Persson et al., 2009 who suggest that Fe in the seeds is mainly associated with PA.

Phytate reduction in a plant is often negatively correlated with its agronomic performance (Bregitzer and Raboy, 2006) which is initially observed as a direct affect on its seed germination ability. We therefore conducted germination trials on our *lpa* transgenics and found that they show a good response, with seedling emergence almost identical to the non-transgenic parent type. Our results seem to be in agreement with the recent reports of Shi et al., 2007 and Spear and Fehr, 2007 who suggest that high levels of PA are not an absolute requirement for efficient seed germination or emergence. Lastly, we studied the morphological traits of each transgenic line and found that the lines generated from events P1 and P4 produced seeds with reduced size and weight. The lower seed yield may be attributed to reduction in starch accumulation resulting from indirect affect of disturbed Pi homeostasis on a key regulatory enzyme in starch biosynthesis pathway (Plaxton and Preiss, 1987).

From our research, we can assert that *GmIPK2* gene is an appropriate candidate for targeted phytate reduction without any pleotropic effects. Despite our results, there is a need to conduct additional field trials under a variety of conditions, across generations and perform their in depth phenotypic and genotypic analysis in order to conclude the effect of *lpa* trait. Further studies are therefore underway in our laboratory in an effort to establish the same.

AUTHOR CONTRIBUTIONS

MP contributed in conception and designing of the study, performed the experiments, acquired the data, performed its analysis and interpretation, and drafted the final manuscript. NB contributed in designing the study, revising the manuscript critically for important intellectual content, and gave final approval of the version to be published. MJ contributed in data acquisition, revising the manuscript critically for important intellectual content, and gave final approval of the version to be published. AD contributed in interpretation of data, revising the manuscript critically for important intellectual content, and gave final approval of the version to be published. AS contributed in conception and designing of the study, revising the manuscript critically for important intellectual content, and gave final approval of the version to be published.

FUNDING

This work was supported by the National Funds for Basic, Strategic and Frontier Application Research in Agriculture (Grant No. NFBSFARA/RNAi -2011/2011-12), ICAR, Government of India.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2018.00804/full#supplementary-material

REFERENCES

Adams, C. L., Hambidge, M., Raboy, V., Dorsch, J. A., Sian, L., Westcott, J. L., et al. (2002). Zinc absorption from a low-phytic acid maize. *Am. J. Clin. Nutr.* 76, 556–559. doi: 10.1093/ajcn/76.3.556

Ali, N., Paul, S., Gayen, D., Sarkar, S. N., Datta, K., and Datta, S. K. (2013a). 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:e68161. doi: 10.1371/journal.pone.0068161

Ali, N., Paul, S., Gayen, D., Sarkar, S. N., Datta, S. K., and Datta, K. (2013b). RNAi mediated down regulation of myo-inositol-3-phosphate synthase to generate low phytate rice. *Rice* 6:12. doi: 10.1186/1939-8433-6-12

Al-Amery, M., Fukushige, H., and Hildebrand, D. (2015). Single seed selection for low phytate lines. *J. Am. Oil Chem. Soc.* 92, 1119–1123. doi: 10.1007/s11746-015-2681-9

Anderson, B. P., and Fehr, W. R. (2008). Seed source affects field emergence of low-phytate soybean lines. *Crop Sci.* 48, 929–932. doi: 10.2135/cropsci2007.09.0510

Bhati, K. K., Alok, A., Kumar, A., Kaur, J., Tiwari, S., and Pandey, A. K. (2016). Silencing of ABCC13 transporter in wheat reveals its involvement in grain development, phytic acid accumulation and lateral root formation. *J. Exp. Bot.* 67, 4379–4389. doi: 10.1093/jxb/erw224

Bregitzer, P., and Raboy, V. (2006). Effects of four independent low-phytate mutations on barley agronomic performance. *Crop Sci.* 46, 1318–1322. doi: 10.2135/cropsci2005.09-0301

Brinch-Pedersen, H., Olesen, A., Rasmussen, S. K., and Holm, P. B. (2000). Generation of transgenic wheat (*Triticum aestivum* L.) for constitutive accumulation of an *Aspergillus* phytase. *Mol. Breed.* 6, 195–206. doi: 10.1023/A:1009690730620

Campion, B., Sparvoli, F., Doria, E., Tagliabue, G., Galasso, I., Fileppi, M., et al. (2009). Isolation and characterisation of an lpa (low phytic acid) mutant in common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 118, 1211–1221. doi: 10.1007/s00122-009-0975-8

Chen, P. S., Toribara, T. Y., and Warner, H. (1956). Microdetermination of phosphorus. *Anal. Chem.* 28, 1756–1758. doi: 10.1021/ac60119a033

Chiera, J. M., Finer, J. J., and Grabau, E. A. (2004). Ectopic expression of a soybean phytase in developing seeds of *Glycine max* to improve phosphorus availability. *Plant Mol. Biol.* 56, 895–904. doi: 10.1007/s11103-004-5293-6

Day, C. D., Lee, E., Kobayashi, J., Holappa, L. D., Albert, H., and Ow, D. W. (2000). Transgene integration into the same chromosome location can produce alleles that express at a predictable level, or alleles that are differentially silenced. *Genes Dev.* 14, 2869–2880. doi: 10.1101/gad.849600

Denbow, D. M., Grabau, E. A., Lacy, G. H., Kornegay, E. T., Russell, D. R., and Umbeck, P. (1998). Soybeans transformed with a fungal phytase gene improve phosphorus availability for broilers. *Poultry Sci.* 77, 878–881. doi: 10.1093/ps/77.6.878

Guttieri, M., Bowen, D., Dorsch, J. A., Raboy, V., and Souza, E. (2004). Identification and characterization of a low phytic acid wheat. *Crop Sci.* 44, 418–424. doi: 10.2135/cropsci2004.4180

Halsted, J. C., Smith, J. C. Jr., and Irwin, M. I. (1974). A conspectus of research on zinc requirements of man. *J. Nutr.* 104, 345–350. doi: 10.1093/jn/104.3.345

Hambidge, K. M., HuVer, J. W., Raboy, V., Grunwald, G. K., Westcott, J. L., Sian, L., et al. (2004). Zinc absorption from low-phytate hybrids of maize and their wild-type isohybrids. *Am. J. Clin. Nutr.* 79, 1053–1059. doi: 10.1093/ajcn/79.6.1053

Hambidge, K. M., Krebs, N. F., Westcott, J. L., Sian, L., Miller, L. V., Peterson, K. L., et al. (2005). Absorption of calcium from tortilla meals prepared from low phytate maize. *Am. J. Clin. Nutr.* 82, 84–87. doi: 10.1093/ajcn/82.1.84

Hegeman, C. E., and Grabau, E. A. (2001). A novel phytase with sequence similarity to purple acid phosphatases is expressed in cotyledons of germinating soybean seedlings. *Plant Physiol.* 126, 1598–1608. doi: 10.1104/pp.126.4.1598

Higgins, T. J. V., Newbigin, E. J., Stransky, H., Llewellyn, D. J., and Craig, S. (1988). The sequence of a pea vicilin gene and its expression in transgenic tobacco plants. *Plant Mol. Biol.* 11, 683–695. doi: 10.1007/BF00017468

Hitz, W., Carlson, T. J., Kerr, P. S., and Sebastian, S. A. (2002). Biochemical and molecular characterization of a mutation that confers a decreased raffinosaccharide and phytic acid phenotype on soybean seeds. *Plant Physiol.* 128, 650–660. doi: 10.1104/pp.010585

Hobbs, S. L., Warkentin, T. D., and Delong, C. M. O. (1993). Transgene copy number can be positively or negatively associated with transgene expression. *Plant Mol. Biol.* 21, 17–26. doi: 10.1007/BF00039614

Jacobsen, T., and Slotfeldt-Ellingsen, K. D. (1983). Phytic acid and metal availability: a study of Ca and Cu binding foodstuffs. *Cereal Chem.* 60, 392–395.

Jiang, S., Cai, W., and Xu, B. (2013). Food quality improvement of soy milk made from short-time germinated soybeans. *Foods* 2, 198–212. doi: 10.3390/foods2020198

Karnwal, M. K., and Singh, K. (2009). Studies on genetic variability, character association and path coefficient for seed yield and its contributing traits in soybean [*Glycine max* (L.) Merrill]. *Legume Res.* 32, 70–73.

Kiers, J. L., Nout, M. J. R., and Rombouts, F. M. (2000). *In vitro* digestibility of processed and fermented soya bean, cowpea and maize. *J. Sci. Food Agric.* 80, 1325–1331. doi: 10.1002/1097-0010(200007)80:9<1325::AID-JSFA648>3.0.CO;2-K

Kuwano, M., Mimura, T., Takaiwa, F., and Yoshida, K. T. (2009). Generation of stable 'low phytic acid' transgenic rice through antisense repression of the 1D-myo-inositol 3-phosphate synthase gene (RINO1) using the 18-kDa oleosin promoter. *Plant Biotechnol. J.* 7, 96–105. doi: 10.1111/j.1467-7652.2008.00375.x

Larson, S. R., Rutger, J. N., Young, K. A., and Raboy, V. (2000). Isolation and genetic mapping of a non-lethal rice (*Oryza sativa* L.) low phytic acid mutation. *Crop Sci.* 40, 1397–1405. doi: 10.2135/cropsci2000.4051397x

Larson, S. R., Young, K. A., Cook, A., Blake, T. K., and Raboy, V. (1998). Linkage mapping of two mutations that reduce phytic acid content of barley grain. *Theor. AppI. Genet.* 97, 141–146. doi: 10.1007/s001220050878

Lehrfeld, J. (1989). High-performance liquid chromatography analysis of phytic acid on a pH-stable, macroporous polymer column. *Cereal Chem.* 66, 510–515.

Li, J., Hegeman, C. E., Hanlon, R. W., Lacy, G. H., Denbow, D. M., and Grabau, E. A. (1997). Secretion of active recombinant phytase from soybean cell-suspension cultures. *Plant Physiol.* 114, 1103–1111. doi: 10.1104/pp.114.3.1103

Linares, L. B., Broomhead, J. N., Guaiume, E. A., Ledoux, D. R., Veum, T. L., and Raboy, V. (2006). Effects of low phytate barley (*Hordeum vulgare* L.) on zinc utilization in young broiler chicks. *Poult. Sci.* 86, 299–308. doi: 10.1093/ps/86.2.299

Liu, Q. L., Xu, X. H., Ren, X. L., Fu, H. W., Wu, D. X., and Shu, Q. Y. (2007). Generation and characterization of low phytic acid germplasm in rice (*Oryza sativa* L.). *Theor. AppI. Genet.* 114, 803–814. doi: 10.1007/s00122-006-0478-9

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Lott, J. N. A. (1984). "Accumulation of seed reserves of phosphorus and other minerals," in *Seed Physiology, Development*, Vol. 1, ed. D. R. Murray (Sydney, NSW: Academic Press), 139–166.

Lott, J. N. A., Greenwood, J. S., and Batten, G. D. (1995). "Mechanisms and regulation of mineral nutrient storage during seed development," in *Seed Development and Germination*, eds J. Kigel and G. Galili (New York, NY: Marcel Dekker), 215–235.

Maupin, L. M., and Rainey, K. M. (2011). Improving emergence of modified phosphorus composition soybeans: genotypes, germplasm, environments, and selection. *Crop Sci.* 51, 1946–1955. doi: 10.2135/cropsci2010.10.0585

Maupin, L. M., Rosso, M. L., and Rainey, K. M. (2011). Environmental effects on soybean with modified phosphorus and sugar composition. *Crop Sci.* 51, 642–650. doi: 10.2135/cropsci2010.07.0396

Mazariegos, M., Hambidge, K. M., Krebs, N. F., Westcott, J. E., Lei, S., Grunwald, G. K., et al. (2006). Zinc absorption in Guatemalan school children fed normal or low-phytate maize. *Am. J. Clin. Nutr.* 83, 59–64. doi: 10.1093/ajcn/83.1.59

Meis, S. J., Fehr, W. R., and Schnebly, S. R. (2003). Seed source effect on field emergence of soybean lines with reduced phytate and raffinose saccharides. *Crop Sci.* 43, 1336–1339. doi: 10.2135/cropsci2003.1336

Mendoza, C., Viteri, F. E., Lonnerdal, B., Young, K. A., Raboy, V., and Brown, K. H. (1998). Effect of genetically modified, low-phytic acid maize on absorption of iron from tortillas. *Am. J. Clin. Nutr.* 68, 1123–1128. doi: 10.1093/ajcn/68.5.1123

Nunes, A., Vianna, G., Cuneo, F., Amaya-Farfán, J., de Capdeville, G., Rech, E., et al. (2006). RNAi-mediated silencing of the myo-inositol-1-phosphate synthase gene (GmMIPS1) in transgenic soybean inhibited seed development and reduced phytate content. *Planta* 224, 125–132. doi: 10.1007/s00425-005-0201-0

Olhoft, P. M., Lin, K., Galbraith, J., Nielsen, N. C., and Somers, D. A. (2001). The role of thiol compounds in increasing *Agrobacterium*-mediated transformation of soybean cotyledonary-node cells. *Plant Cell Rep.* 20, 731–737. doi: 10.1007/s002990100388

Olhoft, P. M., and Somers, D. A. (2001). L-Cysteine increases *Agrobacterium*-mediated T-DNA delivery into soybean cotyledonary node cells. *Plant Cell Rep.* 20, 706–711. doi: 10.1007/s002990100379

Pandey, V., Krishnan, V., Basak, N., Hada, A., Punjabi, M., Jolly, M., et al. (2016). Phytic acid dynamics during seed development and it's composition in yellow and black indian soybean (*Glycine max* L.) genotypes through a modified extraction and HPLC method. *J. Plant Biochem. Biotechnol.* 25, 367–374. doi: 10.1007/s13562-015-0348-0

Paz, M. M., Shou, H. X., Guo, Z. B., Zhang, Z. Y., Banerjee, A. K., and Wang, K. (2004). Assessment of conditions affecting *Agrobacterium*-mediated soybean transformation using the cotyledonary node explant. *Euphytica* 136, 167–179. doi: 10.1023/B:EUPH.0000030670.36730.a4

Persson, D. P., Hansen, T. H., Laursen, K. H., Schjoerring, J. K., and Husted, S. (2009). Simultaneous iron, zinc, sulfur and phosphorus speciation analysis of barley grain tissues using SEC-ICP-MS and IP-ICP-MS. *Metallomics* 1, 418–426. doi: 10.1039/b905688b

Plaxton, W. C., and Preiss, J. (1987). Purification and properties of nonproteolytic degraded ADP glucose pyrophosphorylase from maize endosperm. *Plant Physiol.* 83, 105–112. doi: 10.1104/pp.83.1.105

Prattley, C. A., and Stanley, D. W. (1982). Protein-phytate interactions in soybeans. I. Localization of phytate in protein bodies and globoids. *J. Food Biochem.* 6, 243–253. doi: 10.1111/j.1745-4514.1982.tb00305.x

Raboy, V., Dickinson, D. B., and Below, F. E. (1984). Variation in seed total phosphorus, phytic acid, zinc, calcium, magnesium, and protein among lines of *Glycine max* and *G. soja*. *Crop Sci.* 24, 431–434. doi: 10.2135/cropsci1984.0011183X002400030001x

Raboy, V., Gerbasi, P. F., Young, K. A., Stoneberg, S. D., Pickett, S. G., Bauman, A. T., et al. (2000). Origin and seed phenotype of maize low phytic acid 1–1 and low phytic acid 2–1. *Plant Physiol.* 124, 355–368. doi: 10.1104/pp.124.1.355

Ramteke, R., Kumar, V., Murlidharan, P., and Agarwal, D. K. (2010). Study on genetic variability and traits interrelationship among released soybean varieties of India [*Glycine max* (L.) Merrill]. *Electron. J. Plant Breed.* 1, 1483–1487.

Ramteke, R., and Murlidharan, P. (2012). Characterization of soybean (*Glycine max*) varieties as per DUS guidelines. *Indian J. Agric. Sci.* 82, 572–577.

Rasmussen, S., and Hatzack, F. (1998). Identification of two low-phytate barley (*Hordeum vulgare* L.) grain mutants by TLC and genetic analysis. *Hereditas* 129, 107–112. doi: 10.1111/j.1601-5223.1998.00107.x

Rosche, E., Blackmore, D., Tegeder, M., Richardson, T., Schroeder, H., Higgins, T. J., et al. (2002). Seed-specific overexpression of a potato sucrose transporter increases sucrose uptake and growth rates of developing pea cotyledons. *Plant J.* 30, 165–175. doi: 10.1046/j.1365-313X.2002.01282.x

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Shi, J., Wang, H., Hazebroek, J., Ertl, D. S., and Harp, T. (2005). The maize low-phytic acid 3 encodes a myo-inositol kinase that plays a role in phytic acid biosynthesis in developing seeds. *Plant J.* 42, 708–719. doi: 10.1111/j.1365-313X.2005.02412.x

Shi, J., Wang, H., Schellin, K., Li, B., Faller, M., Stoop, J. M., et al. (2007). Embryo-specific silencing of a transporter reduces phytic acid content of maize and soybean seeds. *Nat. Biotechnol.* 25, 930–937. doi: 10.1038/nbt1322

Spear, J. D., and Fehr, W. R. (2007). Genetic improvement of seedling emergence of soybean lines with low phytate. *Crop Sci.* 47, 1354–1360. doi: 10.2135/cropsci2006.09.0600

Tubana, B. S., and George, T. (1997). *Seed Phosphorus Fertilizer Effect on the Early Growth of Soybean Under Low Phosphorus Supply*. Los Baños: International Rice Research Institute.

Van Leeuwen, W., Ruttink, T., Borst-Vrenssen, A. W., van der Plas, L. H., and van der Krol, A. R. (2001). Characterization of position-induced spatial and temporal regulation of transgene promoter activity in plants. *J. Exp. Bot.* 52, 949–959. doi: 10.1093/jexbot/52.358.949

Veum, T. L., Ledoux, D. R., and Raboy, V. (2007). Low-phytate barley cultivars improve the utilization of phosphorus, calcium, nitrogen, energy, and dry matter in diets fed to young swine. *J. Anim. Sci.* 85, 961–971. doi: 10.2527/jas.2006-453

Wada, T., and Lott, J. N. A. (1997). Light and electron microscopic and energy dispersive x-ray microanalysis studies of globoids in protein bodies of embryo tissues and the aleurone layer of rice (*Oryza sativa* L.) grains. *Can. J. Bot.* 75, 1137–1147. doi: 10.1139/b97-125

Wilcox, J. R., Premachandra, G. S., Young, K. A., and Raboy, V. (2000). Isolation of high seed inorganic P, low-phytate soybean mutants. *Crop Sci.* 40, 1601–1605. doi: 10.2135/cropsci2000.4061601x

Yuan, F. J., Zhao, H. J., Ren, X. L., Zhu, S. L., Fu, X. J., and Shu, Q. Y. (2007). Generation and characterization of two novel low phytate mutations in soybean (*Glycine max* L. Merr.). *Theor. Appl. Genet.* 115, 945–957. doi: 10.1007/s00122-007-0621-2

Yuan, F. J., Zhu, D. H., Deng, B., Fu, X. J., Dong, D. K., Zhu, S. L., et al. (2009). Effects of two low phytic acid mutations on seed quality and nutritional traits in soybean (*Glycine max* L. Merr). *J. Agric. Food Chem.* 57, 3632–3638. doi: 10.1021/jf803862a

Zhang, Z., Xing, A., Staswick, P., and Clemente, T. (1999). The use of glufosinate as a selective agent in *Agrobacterium*-mediated transformation of soybean. *Plant Cell Tissue Org. Cult.* 56, 37–46. doi: 10.1023/A:1006298622969

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.