Identification and Functional Characterization of GmMYB176-Specific Protein Kinases in Soybean

Pravesh Lama

The University of Western Ontario

Supervisor
Sangeeta Dhaubhadel
The University of Western Ontario

Graduate Program in Biology

A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science

© Pravesh Lama 2016

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Biology Commons

Recommended Citation
https://ir.lib.uwo.ca/etd/3823

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact tadam@uwo.ca, wlswadmin@uwo.ca.
Abstract

GmMYB176 is an R1 MYB transcription factor that regulates isoflavonoid biosynthesis in soybean. The deletion of phosphorylation sites in GmMYB176 abrogates its interaction with 14-3-3 proteins and alters its intracellular localization, thus demonstrating the crucial role of phosphorylation in its regulation. While phosphorylation is undoubtedly crucial for GmMYB176 regulation, the identity of kinase(s) responsible remains unknown. Interestingly, a previous LC-MS/MS study using GmMYB176 as bait identified seven putative protein kinases. To further validate these putative GmMYB176-specific protein kinases, the interaction of all seven candidate protein kinases with GmMYB176 were assayed in planta using bimolecular fluorescence complementation assay. All of the candidate protein kinases: Gm02PK, Gm04PK, Gm08PK.1, Gm08PK.2, Gm14PK, Gm17PK and Gm17PPD interacted with GmMYB176 in the nucleus. In addition, Gm17PK also showed interaction with GmMYB176 in the cytoplasm. Furthermore, subcellular localization studies of the individual candidate protein kinases showed that Gm08PK.1 and Gm08PK.2 localize to the nucleus, Gm02PK, Gm04PK and Gm14PK to the endoplasmic reticulum, Gm17PK to the plasma membrane and Gm17PPD to the plastid. The nuclear proteins Gm08PK.1 and Gm08PK.2 were expressed in bacteria and successfully purified for in vitro kinase assays. The results from kinase assays suggested that both Gm08PK.1 and Gm08PK.2 phosphorylate GmMYB176 in vitro. This research will provide the necessary foundation for future studies into the regulation of GmMYB176 and isoflavonoid biosynthesis and thus will help in the metabolic engineering of isoflavonoid biosynthesis in plants.

Keywords

Isoflavonoid, protein kinase, recombinant protein expression, kinase assay
Acknowledgments

Foremost, I would like to express my deepest gratitude to my supervisor Dr. Sangeeta Dhaubhadel who has always provided me guidance and support throughout my Master’s program. I thank my co-supervisor Dr. Jim Karagiannis for his valuable suggestions. I am also indebted to my advisors Dr. Frederic Marsolais and Dr. Anthony Percival-Smith for their useful and constructive recommendations on this project.

I especially thank Ling Chen, Arun Kumaran Anguraj Vadivel and Dr. Jaeju Yu for their abundant help and assistance. I would like to thank my fellow lab mates Arjun Sukumaran, Caroline Sepiol, Hemanta Raj Mainali, Hong Mo (Timothy) Kim, Kishor Duwadi and Dr. Mehran Dastmalchi for being there to share these wonderful two years of time with me.

I would also like to thank Western University for giving me an opportunity to pursue my master degree and London Research and Development Centre for providing laboratory facilities to conduct my research.

Finally, I express gratitude to my family and my friends, who have always believed me and supported me constantly. I will be forever grateful for your love and support.
Table of contents

Abstract ................................................................. i
Acknowledgments ......................................................... ii
Table of contents ........................................................ iii
List of Tables .............................................................. vi
List of Figures .............................................................. vii
List of Appendices ......................................................... ix
List of Abbreviation ...................................................... x

Chapter 1: Introduction ................................................... 1
1.1 Isoflavonoids ......................................................... 1
1.2 Biological activities of isoflavonoids ................................ 1
1.2.1 Human health and nutrition ...................................... 1
1.2.2 Symbiosis and stress resistance in plants ....................... 3
1.3 Regulation of isoflavonoids in soybean ............................ 5
1.4 Protein kinase ......................................................... 7
1.4.1 Protein kinase catalytic domain ................................... 8
1.4.2 Protein phosphorylation mechanism ............................ 10
1.5 Role of kinases and phosphatases in controlling phosphorylated state .... 13
1.6 Identification of GmMYB176-specific protein kinase candidates .......... 15
1.7 Hypothesis .............................................................. 15
1.8 Objectives .............................................................. 15

Chapter 2: Materials and Methods .................................... 17
2.1 Plant materials and growth conditions ................................ 17
2.2 Bacterial strains ....................................................... 17
2.3 *In silico* analysis .................................................................................................................. 17

2.4 Cloning .................................................................................................................................. 18

2.4.1 Cloning into the Gateway entry vector pDONR/Zeo ......................................................... 18

2.4.2 Cloning into destination vectors .......................................................................................... 19

2.5 Transient expression of candidate protein kinases in *N. benthamiana* leaves ............. 25

2.5.1 Subcellular localization of candidate protein kinases ......................................................... 25

2.5.2 *In planta* protein–protein interaction ................................................................................. 25

2.6 Confocal microscopy .............................................................................................................. 26

2.7 Protein expression and purification ....................................................................................... 26

2.8 Tag removal from recombinant protein kinases ................................................................. 27

2.9 *In vitro* Kinase assay ........................................................................................................... 28

**Chapter 3: Results** ................................................................................................................ 30

3.1 *In silico* analysis .................................................................................................................. 30

3.2 Candidate protein kinases contain different kinase domains ............................................ 32

3.3 Conserved residues in the catalytic domain of candidate GmPKs ................................... 32

3.4 Alternate splicing of *Gm02PK* and *Gm08PK* .................................................................. 37

3.5 Candidate GmPKs localize to different compartments of the cell ..................................... 40

3.6 Candidate GmPKs interact with GmMYB176 *in planta* .................................................... 44

3.7 Recombinant Gm08PK.1 and Gm08PK.2 protein expression and purification .............. 47

3.8 Tag removal from recombinant protein kinases ................................................................. 52

3.9 *In vitro* kinase assay .......................................................................................................... 54

3.10 Calculation of free inorganic phosphate generated in the kinase reaction ................. 59

**Chapter 4: Discussion** ........................................................................................................... 61

4.1 Critical residues in the candidate protein kinase catalytic domains ................................. 61
4.2 Candidate GmPKs family and their potential role ........................................ 63
4.3 Subcellular localization of candidate GmPKs ........................................... 65
4.4 Candidate GmPKs interact with GmMYB176 in nucleus ............................. 67
4.5 Non-specific cleavage by enterokinase ..................................................... 69
4.6 Gm08PK.1 and Gm08PK.2 phosphorylate GmMYB176 peptide in vitro .......... 69

Chapter 5: Perspective and future work .............................................................. 72

References ........................................................................................................... 74
Appendices .......................................................................................................... 86
Curriculum Vitae ................................................................................................. 90
List of Tables

Table 2.1 List of primers used for candidate GmPKs amplification............................................. 20
Table 3.1 Characteristics of candidate GmPKs.................................................................................. 31
Table 3.2 Optical density obtained for in vitro kinase assays at 620 nm........................................... 56
Table 4.1 pST binding site for candidate GmPKs ............................................................................. 68
List of Figures

Figure 1.1 Chemical structure of a) isoflavonoid, and b) 17β-estradiol .......................... 2
Figure 1.2 Structure of the catalytic core of a protein kinase. .................................. 9
Figure 1.3 Phosphorylation of a substrate by a protein kinase. .............................. 11
Figure 1.4 Protein regulation through a) phosphorylation and b) de-phosphorylation,
mediated by a protein kinase and a protein phosphatase, respectively .................. 12
Figure 2.1 Map of destination vectors used for expression of candidate GmPKs in N.
benthamiana ........................................................................................................ 21
Figure 2.2 Map of pET32a(+) vector used for protein expression of candidate GmPKs in
E. coli Rosetta(DE3). .............................................................................................. 23
Figure 2.3 Schematic representation of the cloning strategy. .................................. 24
Figure 3.1 Multiple sequence alignment of catalytic domain of candidate GmPKs. ...... 35
Figure 3.2 Schematic representation of the mRNA of Gm02PK derived from its genomic
gene sequence. .................................................................................................... 38
Figure 3.3 Schematic representation of splice variant of Gm08PK.2 ....................... 39
Figure 3.4 Subcellular localization of candidate GmPKs. ......................................... 42
Figure 3.5 GmMYB176 interact with candidate GmPKs in the nucleus .................... 45
Figure 3.6 Optimization of recombinant Gm08PK.1 protein expression in E. coli
Rosetta(DE3) ........................................................................................................ 48
Figure 3.7 Optimization of Gm08PK.2 recombinant protein expression in E. coli
Rosetta(DE3) ........................................................................................................ 49
Figure 3.8 Purification of recombinant protein Gm08PK.1 ........................................... 50
Figure 3.9 Purification of recombinant protein Gm08PK.2 ........................................... 51
Figure 3.10 Enterokinase cleaves recombinant proteins, Gm08PK.1 and Gm08PK.2, at multiple non-specific sites. ................................................................. 53

Figure 3.11 Principle of Universal kinase assay. ................................................................. 55

Figure 3.12 Candidate protein kinase Gm08PK.1 phosphorylates GmMYB176 peptide \textit{in vitro}. .................................................................................................................. 57

Figure 3.13 Candidate protein kinase Gm08PK.2 phosphorylates GmMYB176 peptide \textit{in vitro}. .................................................................................................................. 58

Figure 3.14 A phosphate standard curve generated from known concentrations of inorganic phosphate in a kinase assay buffer................................................................. 60
List of Appendices

Appendix A. Characteristics of candidate protein kinases .......................................................... 86

Appendix B. Multiple sequence alignment of candidate protein kinases catalytic domains .................................................................................................................. 87

Appendix C. Subcellular localization of Gm08PK.1 merged with the confocal image under white light. .............................................................................................................. 88

Appendix D. Gm08PK.1, Gm08PK.2, Gm17PK and Gm17PPD showed interaction with GmMYB176 in reciprocal combination. ................................................................. 89
List of Abbreviation

*Standard SI units not listed

4CL 4-coumarate-CoA-ligase
ADP adenosine diphosphate
ATP adenosine triphosphate
att attachment
BiFC biomolecular fluorescent complementation
bp base pair
cAPK cyclic adenosine monophosphate dependent protein kinase
CDK cyclin-dependent protein kinase
cDNA complementary deoxyribonucleic acid
CFP cyan fluorescent protein
CHI chalcone isomerase
CHR chalcone reductase
CHS chalcone synthase
CPK calcium dependent protein kinase
DNA deoxyribonucleic acid
ER endoplasmic reticulum
h hours
HTH helix-turn-helix
IFS isoflavone synthase
IPTG isopropyl β-1-thigalactopyranoside
ILK Integrin-linked kinase
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>LC-MS/MS</td>
<td>liquid chromatography-tandem mass spectrometry</td>
</tr>
<tr>
<td>LB</td>
<td>lysogeny broth</td>
</tr>
<tr>
<td>MES</td>
<td>2-(N-morpholino)ethanesulfonic acid</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Center for Biotechnology Information</td>
</tr>
<tr>
<td>NLS</td>
<td>nuclear localizing signal</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>PAL</td>
<td>phenylalanine ammonialyase</td>
</tr>
<tr>
<td>PANTER</td>
<td>protein analysis through evolutionary relationships</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>RLK</td>
<td>receptor like protein kinase</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RT</td>
<td>room temperature</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>Ser</td>
<td>serine</td>
</tr>
<tr>
<td>STRADα</td>
<td>Ste20 related adaptor α isoform</td>
</tr>
<tr>
<td>STS</td>
<td>stilbene synthase</td>
</tr>
<tr>
<td>Thr</td>
<td>threonine</td>
</tr>
<tr>
<td>Trx</td>
<td>thioredoxin</td>
</tr>
<tr>
<td>Tyr</td>
<td>tyrosine</td>
</tr>
<tr>
<td>YFP</td>
<td>yellow fluorescent protein</td>
</tr>
<tr>
<td>YN</td>
<td>N-terminal of YFP</td>
</tr>
<tr>
<td>YC</td>
<td>C-terminal of YFP</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

1.1 Isoflavonoids

Isoflavonoids are plant natural products that are produced from the phenylpropanoid pathway. The core chemical structure of isoflavonoids consists of 2 phenyl rings (A and B) and a heterologous C ring (where the B ring is attached to the C-2 carbon of the C ring) (Figure 1.1). They are almost exclusively found in leguminous plants (Lapcik, 2007). Soybean, the most widely grown grain legume, is a rich source of isoflavonoids. Soybean seeds contain three isoflavone aglycones, a subset of isoflavonoids: genistein, daidzein, and glycinein. These aglycones are found in 7-O-β-glucosides, 6"-O-acetyl-β-glucosides and 6"-O-malonly-β-glucosides, respectively (Kudou et al., 1991). Isoflavonoids have become the focus of much recent research due to their many beneficial activities in both humans and plants.

1.2 Biological activities of isoflavonoids

1.2.1 Human health and nutrition

Soybean is an important part of a regular diet in many Asian countries. The consumption of soybean is thought to be related to the low incidence of hormonal-dependent cancers in Asian populations (Watanabe et al., 2002). Many clinical and epidemiological studies have shown that isoflavonoids act as phytoestrogens, plant derived compounds that have an estrogenic effect (Kuiper et al., 1998; Lu et al., 2001). The chemical structure of the isoflavone aglycone, genistein, is similar to the human hormone, 17β-estradiol (Figure 1.1). Genistein can bind to estrogen receptor α, β, and sex hormone binding globulin (Chen and Rogan, 2004b). However, the binding of genistein to the estrogen receptors is weaker as compared to endogenous estrogen. It can elicit either estrogenic or antiestrogenic effects depending on whether the endogenous estrogen concentrations are low, or high, respectively (Wang, 2002; Chen and Rogan, 2004).
Figure 1.1 Chemical structure of a) isoflavonoid, and b) 17β-estradiol
According to Lamartiniere (Lamartiniere, 2002), there is an inverse relationship between cancer susceptibility and mammary gland differentiation. In his experiment, rats administered with genistein during the prepubertal period, or the prepubertal and adult periods, showed differentiated mammary glands and were less susceptible to induced mammary cancer. There are several clinical studies that have demonstrated positive roles for isoflavonoids, or a soy rich diet, in health and nutrition. For example, epidemiological studies in Chinese and Asian American women have shown a positive correlation between high soy intake during adolescence and the reduced risk of breast cancer in adulthood (Shu et al., 2001; Wu et al., 2002). Genistein has also been reported to inhibit prostate cancer cell growth \textit{in vitro}, and prostate tumor development in animals, suggesting that isoflavones may be beneficial to prostate cancer patients (Messina, 2003; Chiyomaru et al., 2013). A study of cancer cell lines showed that daidzein induced cell arrest at G1 phase, whereas genistein arrested the cell cycle at G2/M phase (Matsukawa et al., 1993; Wang et al., 2002). These antitumor activities, independent of estrogenic activity, make isoflavones potential cancer prevention agents (Duffy et al., 2007). Isoflavones may also prevent osteoporosis by binding to estrogen receptors and stimulating bone formation (Watanabe et al., 2002; Pílsková et al., 2010). A clinical trial conducted by the American Heart Association found that cholesterol level decreased after consumption of soy protein, suggesting that soy protein may protect against cardiovascular disease (Erdman and Committee, 2000; Sacks et al., 2006).

1.2.2 Symbiosis and stress resistance in plants

Isoflavonoids play an important role in root nodule formation in legumes (Subramanian et al., 2006). A root nodule is a specialized organ inhabited by nitrogen-fixing bacteria (rhizobia) that fix atmospheric nitrogen into ammonia and provide it to the host-plant. Isoflavonoids released by legume roots act as chemo-attractants for rhizobia and activate the expression of rhizobial \textit{nod} genes (Subramanian et al., 2006). Nod factors, produced from \textit{nod} gene expression, are perceived by legume receptors and trigger a cascade of events. It begins with root hairs curling around invading rhizobia and is followed by the entry of rhizobia into the plant through an infection thread. Nod factors produced from rhizobia then promote the initiation of cell division in the root cortex to form nodule
primordia (Eckardt, 2006). Subramanian et al. (2006) provided the first genetic evidence that isoflavones are essential for soybean root nodule formation. RNA interference (RNAi) induced silencing of *isoflavone synthase* (*IFS*) expression, a key gene involved in isoflavonoid synthesis, resulted in decreased levels of isoflavone and in severely reduced nodulation in soybean hairy root composite plants.

Isoflavonoids also act as phytoalexins. These are low-molecular-weight antimicrobial metabolites that play a role in plant defense against pathogens and injury (Kuc and Rush, 1985; Leifer and Barberio, 2016). Fungal and bacterial extracellular metabolites or their cell wall components are recognized by plant cells and elicit the production of phytoalexins (Kuc, 1984). They accumulate at the site of pathogen infection to inhibit the development of pathogens (Durango et al., 2002). Phytoalexins provide partial resistance against pathogens as they have a low specificity for induction and activity against pathogens (Kuc and Rush, 1985). Glyceollin, one of the extensively studied phytoalexins (Dastmalchi and Dhaubhadel, 2014), inhibits the growth of a wide range of soybean pathogens such as *Phytophthora sojae*, *Sclerotinia sclerotiorum* and *Macrophomina phaseolina* (Lygin et al., 2010). The phytoalexin property of glyceollin was experimentally demonstrated by altering the levels of two isoflavonoid pathway enzymes, chalcone reductase (CHR), and IFS in soybean hairy roots which decreased the levels of glyceollin leading to increased susceptibility to pathogens (Graham et al., 2007).
1.3 Regulation of isoflavonoids in soybean

Isoflavonoid biosynthesis is a complex process that is affected by environmental factors, genetic factors and their interaction (Anguraj Vadivel et al., 2015). A study conducted on two different soybean cultivars, a low isoflavonoid cultivar (Hardin), and a high isoflavonoid cultivar (Corsoy-79), showed an environmental effect on isoflavonoid level in soybean seeds (Eldridge and Kwolek, 1983). The accumulation of isoflavonoids in Hardin and Corsoy-79 grown in Girard, Illinois were 460 and 800 μg/g, respectively. But when grown in Urbana, Illinois Hardin and Corsoy-79 accumulated 820 and 1550 μg of isoflavonoids/g of seed, respectively. Isoflavonoid level was consistently higher in Corsoy-79 seeds compared to Hardin seeds, when grown in the same environment at the same time. Furthermore, a study on global gene expression during embryo development on two different soybean cultivars, a low isoflavonoid cultivar (Harovinton) and a high isoflavonoid cultivar (RCAT-Angora), showed that the cultivar differences had very little effect when compared to either location or growing seasons (Dhaubhadel et al., 2007). Comparison of expression of isoflavonoid-related genes during embryo development in these two cultivars identified two differentially expressed CHS genes (CHS7 and CHS8) that possibly play a critical role in isoflavonoid synthesis in soybean seeds (Dhaubhadel et al., 2007).

There are 9 CHS genes reported in soybean (Tuteja and Vodkin, 2008). Phylogenetic analysis of the soybean CHS gene family has shown that CHS7 and CHS8 form a separate clad in the tree (Yi et al., 2010a). Comparison of CHS7 and CHS8 gene expression during embryo development revealed that CHS8 was expressed from early development to maturity while CHS7 expression was only noticed at the mature stage (Dhaubhadel et al., 2007). It is thought that CHS8 plays a role in isoflavonoid biosynthesis under normal growth conditions while CHS7 plays a role in isoflavonoid production under stress. Furthermore, silencing of the CHS8 gene resulted in a reduction of the transcript levels of both CHS7 and CHS8 and in a reduction of isoflavonoid levels in soybean seed, thus, verifying its role in the regulation of isoflavonoid biosynthesis (Yi et al., 2010a).

CHS genes are regulated by MYB transcription factors in Arabidopsis, maize and several other plants (Sokolova et al., 2012). For example, the isoflavonoid content in soybean seeds
increased when a chimeric transcription factor, CRC from maize, was introduced into soybean (Yu and McGonigle, 2005). The chimeric protein contained an R myc-type transcription factor in between the binding and activation domains of the C1 MYB transcription factor. This led to the speculation that MYB transcription factor(s) might also regulate CHS8 expression in soybean. MYB transcription factors contain several conserved DNA binding domain called MYB domain (Ambawat et al., 2013). The MYB domain can have up to four imperfect repeats (R) (Dubos et al., 2010). Each repeat is made up of 52 amino acid sequences and consists of three α-helices where the second and third helices bind DN. MYB transcription proteins can be divided into 4 groups based upon the number of repeats: R1-MYB, R2R3-MYB, R3-MYB and R4-MYB (Dubos et al., 2010).

In an effort to identify the regulator that controls the expression of CHS8 in soybean, an R1-MYB transcription factor GmMYB176 was identified (Yi et al., 2010b). Furthermore, silencing of GmMYB176 significantly reduced both CHS8 expression and isoflavonoid content in soybean hairy roots, demonstrating that GmMYB176 plays an important role in isoflavonoid synthesis in soybean (Yi et al., 2010b). However, overexpression of GmMYB176 was not sufficient to increase CHS8 expression and isoflavonoid content in soybean hairy roots.

GmMYB176 is predominantly localized to the nucleus despite the fact that it does not contain a nuclear localizing signal (NLS) (Li et al., 2012). The subcellular localization of GmMYB176 is dependent on its phosphorylation state, and its interaction with 14-3-3 proteins (Li et al., 2012). 14-3-3 proteins are the class of phospho-protein binding proteins that bind to phosphoserine/phosphothreonine residues located in their targets, thereby influencing their localization (Yi et al., 2010b). The soybean genome contains 16 functional 14-3-3 proteins and all of them interact with GmMYB176 (Li et al., 2012). The interaction of GmMYB176 with 14-3-3 proteins not only affects the intracellular localization of GmMYB176, but also influences the biosynthesis of isoflavonoids. For example, overexpression of SGF14I in soybean hairy roots decreased isoflavonoid levels (Li et al., 2012). It was also found that the interaction of GmMYB176 with 14-3-3 proteins depends on the phosphorylation of serine-29 position in GmMYB176. The substitution of serine-29 with alanine abolished the interaction between them. Furthermore, the mutant
GmMYB176 S29A was retained in the nucleus only, whereas wild-type GmMYB176 was found in both the nucleus and the cytoplasm, suggesting an important role for the phosphorylation of GmMYB176 at serine-29 (Li et al., 2012). It is speculated that GmMYB176-specific protein kinase(s) phosphorylate GmMYB176 and this phosphorylation leads to its interaction with 14-3-3 proteins and its retention in the cytoplasm. The unphosphorylated fraction goes to the nucleus by some unknown mechanism (Li et al., 2012). The identification of GmMYB176-specific protein kinase(s) will provide insight into the regulation of GmMYB176 and isoflavonoid biosynthesis.

1.4 Protein kinase

Protein kinases are enzymes that regulate signal transduction and are necessary for various aspects of cellular metabolism. They do so by phosphorylation, transferring a phosphate group from a phosphate donor to an amino acid of a substrate. Generally, a protein kinase transfers the γ-phosphate of ATP to its substrate and alters its enzymatic activity (Hunter, 1991). Phosphorylation triggers different responses to extracellular signal molecules such as growth factors and hormones, defense response, cell cycle and response to different stressful environmental conditions (Hardie, 1999).

The protein kinase gene family is one of the largest gene families in eukaryotes, constituting 1-2% of functional genes. Based on the phosphorylation of the OH group of serine (Ser) or threonine (Thr), or the phenolic group of tyrosine (Tyr) in the target protein, protein kinases are classified into two super families: (a) Ser/Thr protein kinase, and (b) Tyr protein kinase (Hanks and Hunter, 1995). Apart from these two superfamilies, a third family of protein kinases, histidine kinases (His), have been shown to phosphorylate the aspartate residue of the target proteins (Hanks and Hunter, 1995). A classification system based on the target amino acid residue creates a very large family of Ser/Thr and Tyr kinase. In order to mitigate this problem, Hanks and Hunter classified protein kinases based upon amino acid sequences within the catalytic domains, the domain responsible for the phospho-transfer reaction (Hunter, 1991). Using a phylogenetic approach, protein kinases were divided into 5 groups, which were further divided into 55 subfamilies. An extended version of Hanks and Hunter classification of protein kinases consists of 3 levels: 9 groups, 81 families and 238 subfamilies (Martin et al., 2010). This classification scheme included
sequence comparison of catalytic domains aided by sequence similarity and domain structure outside the catalytic domain.

1.4.1 Protein kinase catalytic domain

Most eukaryotic protein kinases consist of a catalytic domain tethered to one or more non-kinase domains. The catalytic domains range from 200 to 300 amino acid residues, corresponding to about 30 kDa (Martin et al., 2010). The catalytic domain i) binds and orients ATP or phosphate donor as a complex with the divalent cation ii) binds and orients the substrate and iii) transfers a phosphate to the substrate (Hanks and Hunter, 1995). The non-kinase domain is responsible for regulation, substrate specificity, and scaffoldings (Martin et al., 2010).

Crystal structures of several protein kinases have confirmed the structural conservation of the protein kinase catalytic core (Wang and Cole, 2014). The conservation of the catalytic core implies a conserved mechanism for the transfer of the phosphate group. The first crystal structure of a protein kinase was elucidated for cAPK-α using X-ray diffraction methods that showed the three dimensional arrangement of amino acids (Knighton et al., 1991). The structure revealed a catalytic domain that folded into a two-lobed structure. The smaller amino-terminal (N-terminal) region was predominantly composed of β-sheet structures whereas the larger C-terminal region was composed of α-helices. As shown in Figure1.3, the two lobes formed a cleft in which an ATP molecule could be completely accommodated (Zheng et al., 1993).

Comparing the amino acid residues of protein kinases reveals low and high conservation regions in the catalytic domain (Hanks and Quinn, 1991). The high conservation regions, also termed subdomains, have invariant, or nearly invariant, amino acid residues that play a role in kinase activity (Hanks and Hunter, 1995).
**Figure 1.2 Structure of the catalytic core of a protein kinase.** The kinase domain folds into a two-lobed structure. The N-terminal region is predominantly composed of β-sheet structures whereas the C-terminal is comprised of α-helices. The two domains, linked by a hinge region, form a cleft where ATP (represented by 3D molecular model) binds. Figure adapted from Chico, Van Eldik et al. 2009.
1.4.2 Protein phosphorylation mechanism

Protein phosphorylation is the addition of a phosphate group (PO$_4^{3-}$) to the protein. The phosphate group is transferred from the ATP molecule to the protein (substrate) by an enzyme known as protein kinase. First, an ATP first binds in the cleft between the two lobes of a protein kinase with the adenosine moiety buried in the hydrophobic pocket exposing the γ-phosphate group (Figure 1.4). The substrate protein then binds to the active site of the protein kinase followed by transfer of γ-phosphate to a Ser, Thr or Tyr residue in the substrate. After phosphorylation, the substrate is released from the complex. The ADP is then released from the protein kinase. The order of steps may differ depending on the type of protein kinase. In some cases, the binding of the substrate may precede the binding of ATP and ADP may release before the release of the phosphorylated substrate (Ubersax and Ferrell, 2007).

A transfer of phosphate is initiated by an Asp residue, located in the catalytic domain of a protein kinase (Hanks and Hunter, 1995). It accepts a proton from the hydroxyl (OH) group, present on the Ser, Thr or Tyr residue of the substrate (Lin et al., 2007). The oxygen of the OH group stimulates nucleophilic attack on the γ-phosphate group of ATP, resulting in transfer of the phosphate group. The transfer reaction is facilitated by magnesium (Mg$^{2+}$) that chelates γ- and β-phosphate to lower the threshold for the transfer (Figure 1.4).
**Figure 1.3 Phosphorylation of a substrate by a protein kinase.** a) ATP (red) binds to the cleft of a protein kinase (blue) followed by the binding of a substrate (green). The phosphate group is then transferred to Ser/Thr or Tyr residue in the substrate. Phosphorylated substrate is released before ADP. b) The order of binding/release of substrate and ATP differs for different kinases. Binding of the substrate can precede ATP binding and ADP may be released before phosphorylated substrate. Modified from Ubersax and Ferrell, 2007.
Figure 1.4 Protein regulation through a) phosphorylation and b) de-phosphorylation, mediated by a protein kinase and a protein phosphatase, respectively. Aspartic acid of protein kinase (blue) initiates phosphorylation reaction. It accepts a proton from the hydroxyl group of serine or threonine or tyrosine of a substrate (green). The oxygen of the substrate stimulates nucleophilic attack on the λ-phosphate (phosphate highlighted in red) of an ATP (red), resulting transfer of the λ-phosphate group and an ADP. Two residues of a protein phosphatase (yellow), aspartic acid and cysteine residue initiate phosphate removal from the substrate. Sulphur of the cysteine attacks on the phosphate while the aspartic acid donates a proton to oxygen of substrate and forms a phosphoryl-cysteine intermediate. The negatively charged aspartic acid of the phosphatase then accepts the proton from a water molecule making hydroxyl as a nucleophile. The hydroxyl group then carries out nucleophilic attack on the phosphate group which results breaking of phosphorous-oxygen bond and releasing of phosphate group as hydrogen phosphate. Figure modified from Lin et.al, 2007.
1.5 Role of kinases and phosphatases in controlling phosphorylated state
Phosphorylation is the most common mechanisms to regulate protein functioning and signal transduction. The addition of the negatively charged phosphate group to the protein alters the conformation of the protein. The change in conformation lead to - a) change in catalytic activity of the protein i.e. either activation or inactivation of the protein; b) interactions with other proteins; and c) change in subcellular localization of the protein (Humphreys and Chapple, 2000). This regulation is counterbalanced by phosphatases. They transfer the phosphate group from the phosphorylated substrate to water molecules.

The dephosphorylation reaction is initiated by two residues of the phosphatase. Cysteine of phosphatase stimulates the nucleophilic attack on the substrate phosphate group while aspartic acid of phosphatase donates proton to leaving group oxygen (Figure 1.4) (Tonks, 2006). This forms the phosphoryl-cysteine intermediate. The negatively charged aspartic acid of phosphatase then accepts the proton from a water molecule making hydroxyl (OH) as a nucleophile. The hydroxyl group then carries out nucleophilic attack on the phosphate group which results breaking of phosphorous-oxygen bond and releasing of phosphate group.

Target proteins are phosphorylated at specific sites by one or more protein kinases, and the proteins are dephosphorylated by specific phosphatases. The phosphorylation and dephosphorylation regulates the protein. For example, phosphorylation of glycogen phosphorylase b (inactive), transforms it to active glycogen phosphorylase a form which can break down glycogen to glucose (Johnson, 1992). The dephosphorylation revert back it to inactive form which is unable to catalyze the reaction. Balance between phosphorylation and dephosphorylation is necessary for proper functioning of cell. Imbalance of the phosphor-regulation would result either complete phosphorylation or dephosphorylation of the protein. The balance is maintained by - the concentration of protein kinases and phosphatases; their intrinsic activity; and subcellular localization of both target proteins and their regulators (Hunter, 1995).

Phosphorylation and dephosphorylation do not make a complete picture for regulating proteins. Activators and inhibitors in combination with protein kinases and phosphatases regulate the protein. For example, calmodulin dependent protein kinases are inhibited by
their regulatory domain. In presence of Ca$^{2+}$, calmodulin displaces regulatory domain and phosphorylate itself. The protein kinase now is activated and can phosphorylate its substrate (Hunter, 1995; Purves et al., 2001). Not only proteins are regulated by protein kinases and phosphatases but they themselves could be regulated by protein kinase kinases and phosphatases.

1.6 Identification of GmMYB176-specific protein kinase candidates

In Dr. Dhaubhadel’s laboratory at Agriculture and Agri-Food Canada, a project is being carried out to identify and characterize the GmMYB176 interactome. Translational fusions of GmMYB176 with Yellow fluorescent protein (YFP) at either the N- or C- terminus were created, and expressed in soybean hairy roots. Proteins interacting with GmMYB176-YFP and YFP-GmMYB176 were co-immunoprecipitated using the GmMYB176 fusions as bait. The co-immunoprecipitated proteins were identified by Liquid chromatography–tandem mass spectrometry (LC-MS/MS). Of the many identified interacting proteins, six were protein kinases (Anguraj Vadivel and Dhaubhadel, unpublished).

1.7 Hypothesis

GmMYB176 is a phospho-protein that interacts with 14-3-3 proteins in the nucleus and in the cytoplasm (Li et al., 2012). Based on this, I hypothesize that one (or more) of the six co-immunoprecipitated protein kinases phosphorylate(s) GmMYB176, thereby modulating GmMYB176 localization and isoflavonoid biosynthesis.

1.8 Objectives

The objective of this research is to identify and characterize GmMYB176-specific protein kinase(s) in soybean. This was achieved by:

1. Examining the subcellular localization of the candidate protein kinases.

2. Confirming the interaction of candidate protein kinases with GmMYB176 in planta by bimolecular fluorescence complementation (BiFC).
3. Expressing and purifying the candidate protein kinases in a heterologous system.

4. Performing kinase assays to identify GmMYB176-specific protein kinase(s).
Chapter 2: Materials and Methods

2.1 Plant materials and growth conditions

*Nicotiana benthamiana* seeds were sprinkled on PRO-MIX® BX MYCORRHIZAE™ soil (Rivière-du-Loup) contained in a small tray (5” x 7” x 2”). The tray was kept in a growth room with a 16 h light at 23°C and 8 h dark cycle at 18°C. The relative humidity of 60%-70% and light intensity of 100-150 μmol m⁻² s⁻¹ was maintained in the growth room. After a week, individual seedlings were transferred into new pots and watered regularly. The nutrient mixture of nitrogen, phosphorous, and potassium (20-8-20) was applied three times a week.

2.2 Bacterial strains

*Escherichia coli* strain DH5α (Invitrogen) and *Agrobacterium tumefaciens* strain GV3101 were used for gene cloning and transient expression in *N. benthamiana*, respectively. For maintenance and propagation of pET expression construct *E. coli* strain TOP10 (Invitrogen) was used; for protein expression, *E. coli* strain Rosetta(DE3) (Novagen) was used. For transformation, bacterial cells were electroporated in a Gene Pulser® Cuvette (Bio-Rad Laboratories, Inc.) with a 0.1 cm electrode gap by MicroPulser™ (Bio-Rad Laboratories, Inc.). Electroporation was carried for 4-6 milliseconds at 1.8 kV for *E. coli* strains DH5α, Rosetta(DE3) and Top10, and 2.18 kV for *A. tumefaciens* strain GV3101.

2.3 *In silico* analysis

To obtain detailed information with respect to each candidate protein kinase identified by LC-MS/MS, the Phytozome database ([https://phytozome.jgi.doe.gov/pz/portal.html](https://phytozome.jgi.doe.gov/pz/portal.html)), which contains the whole genome sequence of *Glycine max*, was searched. Based upon the retrieved sequences, PCR primers were designed for the full-length gene amplification of each protein kinase. Protein sizes for the candidate genes and their predicted subcellular locations were determined by using the online tools, Compute pI/Mw ([http://web.expasy.org/compute_pi/](http://web.expasy.org/compute_pi/)) and WoLF-PSORT ([http://www.genscript.com/wolf-psort.html](http://www.genscript.com/wolf-psort.html)), respectively. The kinase domain of candidate protein kinases was studied using PANTHER ([http://www.pantherdb.org](http://www.pantherdb.org)) and National Center for Biotechnology
To determine conserved sequence in catalytic domain candidate GmPKs, the protein sequences of the candidate GmPKs were aligned with characterized Ser/Thr protein kinases using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) followed by BOXSHADE 3.21 (http://ch.embnet.org/software/BOX_form.html). The conserved residues and motifs were manually identified.

2.4 Cloning

2.4.1 Cloning into the Gateway entry vector pDONR/Zeo

Candidate protein kinase genes were amplified using gene-specific primers. The template used in the PCR reaction was cDNA synthesized from RNA isolated from different tissues of soybean. For increasing the specificity in PCR, Gm04PK, Gm08PK and Gm17PPD were amplified using nested PCR. All the primers used for candidate protein kinase gene cloning are listed in Table 2.4.1. Primers for gateway cloning contained attB1 adaptor sequence (5’-GGGGACAAAGTTTGTACAAAAAAGCAGGCT-3’ for forward primer) and attB2 adaptor sequence (5’-GGGG AC CAC TTT GTA CAA GAA AGC TGG GT-3’ for reverse primer). PCR products were separated on a 1% agarose gel and stained with RedSafe (iNtROn Biotechnology). A single band of GmPK amplicon was gel-extracted using EZ-10 Spin Column DNA Gel Extraction Kit (Bio Basic Inc.), quantitated using a NanoDrop spectrophotometer (Thermo Scientific), and recombined into gateway entry vector pDONR/Zeo (Invitrogen) using BP clonase reaction mix (Invitrogen). The BP reaction was transformed into E. coli strain DH5α by electroporation and plated on a lysogeny broth (LB) agar plate containing zeocin (50 μg/mL). E. coli colonies were screened by colony PCR using gene-specific primers. Positive colonies containing pDONR/Zeo-GmPK plasmid was grown overnight at 37°C in 3 mL LB medium containing zeocin (50 μg/mL). Plasmid DNA extraction was performed using the EZ-10 Spin Column Plasmid DNA Kit (Bio Basic Inc.), and sequenced either at the Robarts Research Institute (Western University) or at the London Research and Development Centre, Agriculture and Agri-Food Canada, London, Ontario.
2.4.2 Cloning into destination vectors

After sequences of pDONR/Zeo-GmPKs were confirmed, they were recombined into destination vectors using LR clonase (Invitrogen). For subcellular localization, pEarlyGate101 (Invitrogen) was used and for protein-protein interaction, vectors pEarlyGate201-YN and pEarlyGate202-YC (Invitrogen) were used. The LR reaction mix was transformed into E. coli strain DH5α by electroporation and plated on an LB agar plate containing Kanamycin (50 μg/mL).

For cloning into a protein expression vector pET32a(+), GmPKs were amplified using the primers that contained NcoI and NotI restriction sites in forward and reverse primers, respectively. GmPK amplicon and pET32a(+) vector DNA were digested separately with both NcoI and NotI restriction enzymes at 37°C for 1 h. The digested products were separated on a 1% agarose gel, and gel-extracted using EZ-10 Spin Column DNA Gel Extraction Kit (Bio Basic Inc.). Restriction digested vector DNA and candidate protein kinase gene fragments were mixed in 1:3 molar ratio and their complementary ends were ligated using T4 DNA ligase (Invitrogen) at 16°C overnight. The ligation reaction mix (1 μL) was transformed into E. coli strain TOP10 cells by electroporation and plated on an LB agar plate containing ampicillin (50 μg/mL).

Positive colonies in the destination vectors were screened using colony PCR with gene-specific primers. Plasmid DNA was extracted using the EZ-10 Spin Column Plasmid DNA Gel Extraction Kit (Bio Basic Inc.) and their sequence were confirmed. For protein-protein interaction and subcellular localization, the destination vectors containing candidate protein kinase genes were transformed into A. tumefaciens strain GV3101, and plated on an LB agar plate containing rifampicin (10 μg/mL), gentamycin (50 μg/mL) and kanamycin (50 μg/mL). For protein expression in bacteria, pET32a(+)GmPKs were transformed into E. coli strain Rosetta(DE3) and plated on an LB agar plate containing ampicillin (50 μg/mL) and chloramphenicol (32 μg/mL). The transformed E. coli colonies were screened by colony PCR using gene-specific primers.
<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5’-3’)</th>
<th>Primer annealing temperature (°C)</th>
<th>Amplicon size (base pairs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PKGm02F1</td>
<td>GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGCTCACAGA6CAACTTACTT</td>
<td>52</td>
<td>2172</td>
</tr>
<tr>
<td>Gm02PKR</td>
<td>GGGGACCACTTTTGCTACAAGAAGGCTGGTCTTTATCTATATTGCAA CTCCTTTCAGAAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gm04PKF1</td>
<td>TTTTGCGTTGAGGTGGATCG</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>Gm04PKR1</td>
<td>TCCATAAAATGCGCCGACAGT</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>Gm04PKF</td>
<td>GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGCTCACAGAACCACATGCTATTTTC</td>
<td>55</td>
<td>2784</td>
</tr>
<tr>
<td>PKGm04R</td>
<td>GGGGACCACTTTTGCTACAAGAAGGCTGGTCTGAATCTGCAATCCGCTCAATCCTGCT</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>PKGm08F1</td>
<td>CCGATCTCTCGCAGCTTCTT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PKGm08R1</td>
<td>CTCTGTCAAACAAACCAAGTTCCTCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gm08PKF</td>
<td>GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGCGGCTAGCAGACATGGCGG</td>
<td>55</td>
<td>2238</td>
</tr>
<tr>
<td>Gm08PKR</td>
<td>GGGGACCACTTTTGCTACAAGAAGGCTGGTCTCAAGGCTCAATCAAATCAGCCGCTCAATCCTGCT</td>
<td>55</td>
<td>2238</td>
</tr>
<tr>
<td>Gm08PKFr</td>
<td>ATATCCCTGCTATGCGCTAGACAGACATGGCGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gm08PK.1Rr</td>
<td>ATATCGCTGCGCGCCTTCCCACCAGCTCCTGTTTT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gm08PK.2R</td>
<td>GGGGACCACTTTTGCTACAAGAAGGCTGGTCTCCGACGAACCAC</td>
<td>55</td>
<td>1962</td>
</tr>
<tr>
<td>Gm08PK.2Rr</td>
<td>ATATGCCCGGCCGCGAGAGATTGTACAGACCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gm014PKF1</td>
<td>TCTTCCACTTGAAGCGAGCGAGGC</td>
<td>55</td>
<td>2106</td>
</tr>
<tr>
<td>Gm014PKR1</td>
<td>CATTTCCACTTGAAGCGAGCGAGGC</td>
<td>55</td>
<td>2106</td>
</tr>
<tr>
<td>PK17FattB1</td>
<td>1 GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAGAATGGTTGTCAAGATGCC</td>
<td>55</td>
<td>1617</td>
</tr>
<tr>
<td>PK17RattB1</td>
<td>2 GGGGACCACTTTTGCTACAAGAAGGCTGGTCAAGAAGGCTCCC TGGTGG</td>
<td>55</td>
<td>1617</td>
</tr>
<tr>
<td>Gm17PPDF</td>
<td>ACATAGAGAACACAGGGAGGAAAAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gm17PPDF1</td>
<td>GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAGAATGGTTGTCAAGATGCC</td>
<td>55</td>
<td>1617</td>
</tr>
<tr>
<td>Gm17PPDR</td>
<td>GGGGACCACTTTTGCTACAAGAAGGCTGGTCAAGAAGGCTCCC TGGTGG</td>
<td>55</td>
<td>1617</td>
</tr>
</tbody>
</table>

*attB1* and *attB2* adapter sites are underlined, *Nco*I and *Not*I restriction sites are shown in bold and italics.
Figure 2.1 Map of destination vectors used for expression of candidate *GmPKs* in *N. benthamiana*. a) pEarlyGate101 was used for subcellular localization studies. It contains YFP translationally fused with the gene of interest at the C-terminus end. b) pEarlyGate201-YN and c) pEarlyGate202-YC were used for BiFC assays to study protein-protein interaction. These vectors contain split YFP: pEarlyGate201-YN contains N-terminal amino acids of YFP and pEarlyGate202-YC contains C-terminal half of YFP. All destination vectors contain a ccdB cassette flanked by two recombination sites attR1 and attR2 for successful recombination events. Destination vectors have the cauliflower mosaic virus 35S promoter to drive the expression of genes in the vector; OCS, the 3’ sequences of the octopine synthase gene for transcription termination; BAR, the Basta herbicide resistance gene for selection of transgenic plants; Km, the bacterial kanamycin resistance gene for bacterial selection; and left border (LB) and right border (RB) sequences for Agrobacterium-mediated T-DNA transfer.
a) pEarlyGate101
12453 bp

b) pEarlyGate201-YN
12321 bp

c) pEarlyGate202-YC
12027 bp
Figure 2.2 Map of pET32a(+) vector used for protein expression of candidate GmPKs in *E. coli* Rosetta(DE3). pET32a(+) contains T7 inducible promoter that is induced by IPTG. Candidate GmPKs are inserted at *Not*I and *Nco*I, restriction sites. Recombinant protein kinases are fused with Trx tags, His tags and S tags that can be cleaved at Ent by Enterokinase protease. The vector has *lacI* gene that encodes lac repressor which turns off recombinant protein expression unless IPTG is introduced. The vector has ampicillin resistance gene for bacterial selection.
Figure 2.3 Schematic representation of the cloning strategy. GmPK amplicons are recombined into entry vector pDONR/Zeo by BP recombination. Through LR recombination, GmPKs from pDONR/Zeo-GmPK are then recombined into pEarlyGate101 for subcellular localization study; and pEarlyGate201-YN and pEarlyGate202-YC to study protein-protein interaction with GmMYB176 in planta. pET32a(+) vectors are digested with NotI and NcoI restriction enzymes and GmPKs are then ligated into pET32a(+). GmPKs are cloned into pET32a(+) for protein expression.
2.5 Transient expression of candidate protein kinases in \textit{N. benthamiana} leaves

2.5.1 Subcellular localization of candidate protein kinases

To study the subcellular localization of candidate GmPKs, \textit{N. benthamiana} leaves were infiltrated with \textit{A. tumefaciens} GV3101 harboring pEarlyGate101-GmPK (Sparkes et al., 2006). Cloning GmPKs into pEarlyGate101 resulted in the translational fusion of each gene with YFP. A single \textit{A. tumefaciens} GV3101 harboring pEarlyGate101-GmPK colony was cultured in infiltration culture medium (LB containing 10 mM 2-N-morpholino-ethanesulfonic acid [MES] pH 5.6, and 100 μM acetosyringone) supplemented with kanamycin (50 μg/mL), rifampicin (10 μg/mL), and gentamycin (50 μg/mL). The cultures were grown at 28°C until the OD\textsubscript{600} reached 0.5-0.8. The cultures were centrifuged at 3000 rpm (Thermo Scientific Sorvall RC 6 Plus) for 30 min. The pellets were resuspended in Gamborg’s solution (3.2 g/L Gamborg’s B5 and vitamins, 20 g/L sucrose, 10 mM MES pH 5.6, and 200 μM acetosyringone), making its OD600 to 1, and incubated with gentle agitation for 1 h. To confirm the subcellular localization of a protein kinase, \textit{A. tumefaciens} GV3101 suspension culture carrying pEarlyGate101-GmPK was co-infiltrated with the respective organelle marker in 1:1 ratio on the ventral side of 4-6 weeks old \textit{N. benthamiana} leaves. Two days after infiltration, the subcellular localization of each candidate protein kinase was studied using confocal microscopy.

2.5.2 \textit{In planta} protein-protein interaction

To confirm the interaction of candidate protein kinases (identified by co-immunoprecipitation and LC-MS/MS) with GmMYB176, bimolecular fluorescence complementation assay (BiFC) was carried out. In this method, split fluorescent protein segments are attached to two interacting proteins. As a result of their interaction, the two segments of the fluorescent protein come in close proximity to form a functional fluorophore that can be detected by confocal microscopy (Kerppola, 2006).

Translational fusion of each candidate GmPK was constructed in pEarlyGate201-YN or pEarlyGate202-YC that contained the N-terminal (YN) or C-terminal (YCYC) half of yellow fluorescent protein (YFP), respectively. Translational fusions of GmMYB176 with both
YN and YC were available in the Dhaubhadel lab (Li et al., 2012). A single colony of *A. tumefaciens* GV3101 harboring GmPK with one of the YFP split or GmMYB176 with the other half of the YFP split, was cultured in infiltration culture medium and processed as described in section 2.5.1. To study the protein-protein interaction, *A. tumefaciens* suspension cultures, containing pEarlyGate201-GmPK-YN or pEarlyGate202-GmMYB176-YC or the reciprocal combination, pEarlyGate201-GmMYB176-YN or pEarlyGate202-GmPK-YC, were mixed in 1:1 ratio and infiltrated into ventral side of *N. benthamiana* leaves using 1 mL needle-less syringe. After infiltration, plants were returned back to growth room. Two days after infiltration, protein expression and interaction were studied by confocal microscopy.

2.6 Confocal microscopy

A Leica TCS SP2 inverted, confocal microscope was used to study both protein-protein interactions between GmPKs and GmMYB176, and investigate the subcellular localization of candidate GmPKs. Epidermal cell layers of *N. benthamiana* leaves were visualized using a 63X water immersion objective lens at an excitation wavelength of 514 nm and 434 nm for YFP and cyan fluorescent protein (CFP), respectively. The emissions were collected between 530-560 nm wavelength for YFP and 460-490 nm for CFP. To record the fluorescence images from two different fluorophores (YFP and CFP) in sequential order as in the case of co-infiltration, ‘Sequential Scan Tool’ was used.

2.7 Protein expression and purification

To obtain purified protein, which could be used for kinase assay, candidate GmPKs were expressed in a bacterial system. A single colony of *E. coli* strain Rosetta(DE3) containing pET32a(+)-Gm08PK.1/Gm08PK.2 was cultured in 3 mL LB medium containing ampicillin (50 µg/mL) and chloramphenicol (32 µg/mL) at 37°C overnight. One mL of overnight grown bacterial cultures was then transferred to 400 mL of LB medium containing ampicillin (50 µg/mL) and chloramphenicol (32 µg/mL) supplemented with 1% glucose at 37°C. The cultures were grown until an OD<sub>600</sub> of 0.5 was reached. Recombinant protein Gm08PK.1 was induced with 0.3 mM IPTG at 18°C for 24 h whereas recombinant Gm08PK.2 was induced with 0.3 mM IPTG at 25°C for 11 h. Cells were harvested by
centrifugation at 5000 rpm using SLA1500 centrifuge (Thermo Scientific Sorvall RC 6 Plus) for 5 min. Pellets were incubated in 50 mL lysis buffer containing 1 mM PMSF, 300 mM NaCl and 100 μg/mL lysozyme for 30 min at 4°C followed by French press (American Instrument Co., Inc., Silver Spring, Md.) at 18000 psi. The cell lysates were sonicated using Qsonica Sonicator Q125 (Fisher Scientific) at 40% amplitude for 1 min with 10 s intervals after 10 s pulse. The cell lysates were centrifuged at 12000 rpm using SLA1500 centrifuge (Thermo Scientific Sorvall RC 6 Plus) for 15 min at 4°C and supernatants were collected and filtered through 0.45 µm filter (SARSTEDT).

To purify recombinant Gm08PK.1/Gm08PK.2 protein, 1 mL HiTrap HP columns (GE Healthcare) prepacked with Ni Sepharose High Performance were charged with 0.1 M NiSO₄, followed by column equilibration with 5 mL binding buffer containing 0.02 M sodium phosphate, 0.5 M NaCl at pH 7.4. Filtered supernatant fractions of crude bacterial cell lysate were then passed through the equilibrated columns. The flow rate was maintained at 1 mL/min while supernatants were passed through the column. The columns were then washed with 5 mL of binding buffer followed by elution with 5 mL of five different concentrations (50 mM, 100 mM, 150 mM, 200 mM and 500 mM) of imidazole in the elution buffer (0.02 M sodium phosphate, 0.5 M NaCl) separately. The eluates were collected in 1 mL of fractions at 1 mL/min flow rate. Eluates (15 μL) from each fraction were separated on a 8% SDS-polyacrylamide gel and stained with Coomassie Blue. The concentration of purified recombinant was determined by Bradford assay (Bradford, 1976).

2.8 Tag removal from recombinant protein kinases

In order to obtain candidate protein kinase in its native form, N-terminal Trx tag, S tag and His tag are cleaved from the recombinant proteins using protease, enterokinase (New England Biolabs). The cleavage reaction was carried out in a buffer containing 25 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), 150 mM NaCl, 10 mM MgCl₂ and 10 mM CaCl₂, pH 7 (kinase assay buffer). 10 μg of recombinant Gm08PK.1/Gm08PK.2 was incubated with 0.2 μg of enterokinase in a kinase assay buffer at 18°C (for 16 h), 25°C (for 5 h, 10 h, and 15 h) and 37°C (for 2 h, 4 h, and 6 h). The
reactions were stopped by heating for 10 min at 95°C. 20 µL of the reaction was separated on a 8% SDS-polyacrylamide gel and stained with Coomassie Blue.

2.9 In vitro Kinase assay

To identify GmMYB176-specific protein kinase, in vitro kinase assays were performed with purified recombinant protein Gm08PK.1/Gm08PK.2, in conjunction with GmMYB176 peptide. The fourteen amino acid peptide of GmMYB176, SMRKSVSMNNLSQY contains serine-29 (underlined) that is phosphorylated and is responsible for interaction with 14-3-3 proteins in soybean. The GmMYB176 peptides were purchased from Bio Basic Inc. and contained more than 95% purity. Peptides were dissolved in water followed by addition of acetic acid (10% of final volume). For in vitro kinase assays, the Universal Kinase Activity Kit (R&D SYSTEMS) was used.

In vitro kinase assays were performed in 96-well microplates in a 50 µL kinase reaction. Kinase reactions were performed using 0.2 µg of recombinant Gm08PK.1 or 0.12 µg of recombinant Gm08PK.2, 0.18 mM GmMYB176 peptide substrate (SMRKSVSMNNLSQY), 0.2 mM ATP and 0.1 µg of coupling phosphatase, CD39L2. The candidate protein kinases, peptide, ATP and coupling phosphatase were prepared in kinase assay buffer. The reactions were incubated at room temperature (RT) for 10 min and terminated by addition of Malachite Green reagent A (30 µL). Malachite Green reagent B (30 µL) was added and incubated for 20 min at RT to stabilize color development. The absorbance of the color was measured at 620 nm using Synergy™ 2 microplate reader (BioTek). For positive control, ADP was used instead of ATP, in the kinase reaction. For negative control, recombinant protein Gm08PK.1/Gm08PK.2 was substituted with kinase assay buffer in the kinase reaction.

Kinase assays for each candidate protein kinase were carried out in three biological replicates. Each biological replicate was obtained from the average of two technical replicates. The mean of three biological replicates was calculated along with its standard error. In order to measure the total amount of phosphate released in the kinase reaction, optical density was correlated with the standard curve. The standard curve was plotted by
measuring optical density (620 nm) at different amount of phosphate (0 pmol to 5000 pmol).
Chapter 3: Results

GmMYB176 is an R1 MYB transcription factor that regulates CHS8 expression and affects isoflavonoid biosynthesis in soybean (Yi et al., 2010b). It interacts with 14-3-3 proteins, a group of phospho-protein binding proteins, that alter its intracellular localization (Li et al., 2012). Wild-type GmMYB176 is found in both the nucleus and in the cytoplasm. In contrast, site-directed GmMYB176 mutants - in which serine-29 is converted to alanine - fail to interact with 14-3-3 proteins and are found exclusively in the nucleus. This clearly demonstrates that the subcellular localization of GmMYB176 is affected through phosphorylation by an as yet unidentified GmMYB176-specific protein kinase.

3.1 In silico analysis

In an effort to identify GmMYB176 protein complex, proteins interacting with GmMYB176 were co-immunoprecipitated and identified by LC-MS/MS. Along with many other interactors, six protein kinases were identified (Anguraj Vadivel and Dhaubhadel, unpublished). To obtain more detailed information with respect to these six candidate protein kinases, the soybean genome database (https://phytozome.jgi.doe.gov/pz/portal.html) was searched and their full coding sequences were obtained. The predicted protein size of the candidate GmPKs ranged from 59 to 104 kDa. The candidates, Gm02PK, Gm04PK, and Gm08PK were predicted to localize to the nucleus. The other candidate protein kinases Gm14PK and Gm17PPD were predicted to localize to the plastid while Gm17PK was predicted to localize to the cytoplasm. Detailed characteristics of each candidate protein kinase are shown in Table 3.1.
Table 3.1 Characteristics of candidate *GmPKs*

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Locus name</th>
<th>Coding sequence (nt)</th>
<th>Predicted protein molecular weight (kDa)</th>
<th>Predicted Subcellular localization</th>
<th>Predicted protein kinase classification</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Gm02PK</em></td>
<td>Glyma.02G215300</td>
<td>2793</td>
<td>104.11</td>
<td>Nucleus</td>
<td>Ethylene response protein kinase CTR1</td>
</tr>
<tr>
<td><em>Gm04PK</em></td>
<td>Glyma.04G096000</td>
<td>2784</td>
<td>103.51</td>
<td>Nucleus</td>
<td>Ethylene response protein kinase CTR1</td>
</tr>
<tr>
<td><em>Gm08PK</em></td>
<td>Glyma.08G102600</td>
<td>2238</td>
<td>87.7</td>
<td>Nucleus</td>
<td>Cyclin-dependent protein kinase</td>
</tr>
<tr>
<td><em>Gm14PK</em></td>
<td>Glyma.14G118800</td>
<td>2106</td>
<td>80.86</td>
<td>Plastid</td>
<td>Receptor like protein kinase</td>
</tr>
<tr>
<td><em>Gm17PK</em></td>
<td>Glyma.17G013800</td>
<td>1617</td>
<td>59.77</td>
<td>Cytoplasm</td>
<td>Calcium dependent protein kinase</td>
</tr>
<tr>
<td><em>Gm17PPD</em></td>
<td>Glyma.17G020600</td>
<td>2853</td>
<td>104.28</td>
<td>Plastid</td>
<td>Pyruvate phosphate dikinase</td>
</tr>
</tbody>
</table>

nt: nucleotide; kDa: kilo-Dalton.
3.2 Candidate protein kinases contain different kinase domains

The catalytic domain of a protein kinase can reveal its substrate specificity and mode of regulation (Hanks and Hunter, 1995). Thus, characterizing the candidate protein kinase catalytic domains may help to predict GmMYB176-specific protein kinase(s). A domain search was performed for all the candidate protein kinases. The results revealed that candidate protein kinases can be categorized into five different families. Gm02PK and Gm04PK are predicted to belong to the family of ethylene response protein kinases that includes CTR1. The other candidates, Gm08PK, Gm14PK, Gm17PK and Gm17PPD were possessed catalytic domains similar to cyclin-dependent protein kinases (CDKs), receptor like protein kinases (RLKs), calcium dependent protein kinases (CPKs) and pyruvate phosphate dikinases (PPDs), respectively.

3.3 Conserved residues in the catalytic domain of candidate GmPKs

A protein kinase catalytic domain contains conserved critical residues that are considered important for kinase activity (Hanks and Hunter, 1995). Therefore, the conserved residues in the candidate GmPKs were identified through alignment with cyclic adenosine monophosphate dependent protein kinase (cAPK-α) and Phaseolus vulgaris protein kinase, PVPK-1. cAPK-α is included because it is the most widely studied protein kinase and is regarded as the universal structural reference for the catalytic domains of protein kinases (Bossemeyer, 1995). PVPK-1 is a Ser/Thr protein kinase from P. vulgaris (a member of Leguminosae family) and contains all the expected conserved residues in its catalytic domain (Lawton et al., 1989). Protein sequences from Gm02PK, Gm04PK, Gm08PK, Gm14PK and Gm17PK were aligned with characterized Ser/Thr protein kinases. Candidate Gm17PPD was excluded in this alignment as it is predicted to be a pyruvate orthophosphate dikinase.

Amino acids that are highly conserved in the catalytic domain of protein kinases fall into 12 different subdomains of the catalytic domain. The 12 subdomains represent localized regions of high conservation and help the catalytic domain fold into a conserved two-lobed structure (Hanks and Quinn, 1991). Conserved amino acids in candidate GmPKs are indicated in Figure 3.1.
The invariant or nearly invariant residues in subdomain I are Gly50 and Gly52. Gly50 is conserved in Gm02PK, Gm04PK, Gm14PK, and Gm17PK, whereas Gly52 is conserved in all five aligned candidate GmPKs. In addition, Gly55, which is more variable among Ser/Thr and Tyr kinases (Hemmer et al., 1997), is conserved in all candidate GmPKs. These residues help to anchor ATP in the cleft (Hanks and Hunter, 1995). The invariant residue, Lys72, in subdomain II, is conserved in all five aligned candidate GmPKs. It interacts with β- and γ-phosphates of ATP and helps to anchor and orient ATP (Zheng et al., 1993). Glu91, in subdomain III, forms a salt bridge with Lys72 and helps it to anchor ATP (Hanks and Hunter, 1995). It is conserved in Gm02PK, Gm04PK, Gm08PK and Gm17PK.

The critical residues Asp166, Lys168 and Asn171 in subdomain VIB are conserved in Gm04PK, Gm08PK, Gm14PK and Gm17PPD, but not in Gm02PK. The Asp166 residue accepts the proton from the hydroxyl group of the substrate amino acid during phosphoryl transfer. Lys168 is another critical residue which interacts with the γ-phosphate of ATP and stabilizes the transition state. Asn171 chelates the secondary Mg$^{2+}$ ion (Hanks and Hunter, 1995). Asp184, in subdomain VII, is conserved in all four aligned candidate GmPKs except Gm02PK. It chelates the primary Mg$^{2+}$ ion that bridges the β- and γ-phosphates (Zheng et al., 1993).

The other important residues - Gly186 in subdomain VII; Glu208 in subdomain VIII; Asp220 and Gly225 in subdomain IX; and Arg280 in subdomain XI - are conserved in Gm04PK, Gm08PK, Gm14PK and Gm17PK. Although the role of these residues in phosphorylation is not known, it is predicted that they help to form the conserved structure of catalytic domain (Hanks and Hunter, 1995).

Apart from critical residues, the Gly-Thr/Ser-X-X-Tyr/Phe-X-Ala-Pro-Glu motif in VIII subdomain (that is specific to Ser/Thr protein kinases) is conserved in Gm04PK, Gm14PK, and Gm17PK. As GmMYB176 is phosphorylated at serine-29, regulatory GmMYB176-specific protein kinase must be Ser/Thr protein kinases. This Ser/Thr specific sequence lies in an activation loop. An activation loop extends from the DFG to the APE motif and
contains Ser/Thr as a phosphorylation site (Nolen et al., 2004). Generally, a protein kinase is regulated by phosphorylation on the activation loop (Nolen et al., 2004).
Figure 3.1 Multiple sequence alignment of catalytic domain of candidate GmPKs. Candidate GmPKs with Ser/Thr kinases-cAPK-α (Human) and PVPK-1 (*Phaseolus vulgaris*) were aligned using Clustal Omega, and imported into BOXSHADE 3.21 for shading. Identical and similar amino acids are shown in dark and light grey boxes, respectively. Important residues that play a role in kinase activity are colored blue; critical residues whose absence will either result in a loss of kinase activity or severe decrease in kinase activity are colored red; and important motifs are underlined. The Gly-Thr/Ser-X-X-Tyr/Phe-X-Ala-Pro-Glu motif (underlined with green) is a Ser/Thr protein kinase specific motif. Roman numerals at the bottom refer to subdomains across the protein kinase family. The numbers above the colored amino acid residue represent the position of that amino acid in cAPK-α. The number preceding the amino acid sequence represents the position of amino acid for the given protein kinase. Gaps are represented by dashes to optimize the alignment. The regions of the sequence that are not conserved are not shown in the figure.
3.4 Alternate splicing of *Gm02PK* and *Gm08PK*

To confirm the interaction between the candidate GmPKs and GmMYB176, each of the candidate genes was amplified using cDNA synthesized from RNA isolated from soybean cultivar Harosoy 63 and cloned into the protein expression vectors (section 2.6). A complete sequence match was found for *Gm04PK, Gm14PK, Gm17PK* and *Gm17PPD* with the sequences deposited in the soybean genome database (https://phytozome.jgi.doe.gov/pz/portal.html). The gene loci are shown in Table 3.1. However, my results show that *Gm02PK* (Glyma.02G215300) and *Gm08PK* (Glyma.08G102600) do not match perfectly with their respective loci.

The coding sequence of Glyma.02G215300 is predicted to be comprised of 2193 bp, encoding a protein of 730 amino acid residues. However, the open reading frame of *Gm02PK* was comprised of only 2172 bp. This was confirmed by the complete sequencing of *Gm02PK* in five different clones. The comparison of the coding sequence from the soybean genome database and the clone revealed a deletion of a 12 bp segment, extending from nucleotide 1530 in exon 5, and a 21 bp deletion, extending from nucleotide 2160 in exon 8 (Figure 3.2). In contrast, 16 bp insertion, 224 nucleotides downstream of the coding sequence, was incorporated into the open reading frame of *Gm02PK*.

In the case of *Gm08PK*, two transcripts were detected using the same pair of *Gm08PK*-specific primers. Complete sequencing of the clones containing *Gm08PK* revealed that in addition to the sequence identical to Glyma.08G102600, a splice variant of the gene was also found. The primary transcript (Glyma.08G102600) was named *Gm08PK.1* and its splice variant was named *Gm08PK.2*. As shown in Figure 3.3, a 25 bp sequence from an intron was inserted into *Gm08PK.2*. This insertion introduced a premature stop codon at nucleotide 1960. This was confirmed by the complete sequencing of *Gm08PK.2* in five different clones. A search for *Gm08PK.2* like sequences in the NCBI database identified a predicted protein with 100% amino acid identity from *Glycine max* (accession number XP_014634338.1).
Figure 3.2 Schematic representation of the mRNA of *Gm02PK* derived from its genomic gene sequence. The genomic sequence of Glyma.02G215300 is represented by blue boxes (exons) and black lines (introns). Compared to the predicted sequence, the actual mRNA contains a 12 bp deletion in exon 5, starting from nucleotide 1530; 21 bp deletion in exon 8, starting at nucleotide 2160 bp; and an insertion of 16 bp after exon 8.
Figure 3.3 Schematic representation of splice variant of *Gm08PK.2*. The genomic sequence of Glyma.08102600 is represented by blue boxes (exons) and black lines (introns). In addition to the primary transcript, a splice variant exists in which a 25 bp insertion is incorporated into the mRNA. It shifts the frame and introduces a premature stop codon at nucleotide 1960.
I incorporated splice variant of Gm08PK.2 as a candidate GmMYB176-specific protein kinase. Like other candidate protein kinases, Gm08PK.2 characteristic features were also obtained. Gm08PK.2 is predicted to be 73.05 kDa, to localize to nucleus, and to contain a cyclin-dependent kinase domain (Appendix A). Also, Gm08PK.2 was aligned with other candidate protein kinases along with the characterized protein kinases to determine the presence of critical residues that are essential for kinase activity (Appendix B). Gm08PK.1 contains all the important amino acid residues mentioned in section 3.3. In the case of Gm08PK.2, it doesn’t contain one of the conserved residues – Arg280. The lack of Arg280 might decrease the kinase activity of Gm08PK.2.

3.5 Candidate GmPKs localize to different compartments of the cell

GmMYB176 is a transcription factor, and is found both in the nucleus and in the cytoplasm (Li et al., 2012). It is speculated that GmMYB176-specific protein kinase(s) localize either in the cytoplasm or in the nucleus. Thus, the localization of GmPKs was studied.

To determine the subcellular localization of candidate GmPKs, each candidate was translationally fused with YFP. GmPK-YFP fusion proteins were transiently expressed in the leaves of N. benthamiana and protein localization were monitored by confocal microscopy. The results indicated that GmPKs localize to different compartments of the cell. To confirm the localization of each GmPK, CFP targeted to specific organelles was used as a marker. They were co-infiltrated with GmPK-YFP into N. benthamiana leaves followed by confocal microscopy.

ER targeted proteins display net-like structure (English and Voeltz, 2013; Nziengui et al., 2007) as shown by Gm02PK, Gm04PK and Gm14PK in Figure 3.4 A, B and E. Although Gm02PK, Gm04PK and Gm14PK were predicted nuclear proteins, they were found to localize to ER. The subcellular localization was also confirmed by overlapping signals from ER targeted CFP and GmPK-YFP.

Both Gm08PK.1 and Gm08PK.2 localize to the nucleus. To confirm the nuclear localization, co-expression of GmPK-YFP fusion protein with nuclear localization signal-
CFP (NLS-CFP) fusion protein was performed. The overlap of signals from Gm08PK.1/Gm08PK.2-YFP and NLS-CFP confirmed their nuclear localization (Figure 3.4 C, D). For Gm08PK.1, the overlapping signals were merged with the confocal image taken under white light (Appendix C).

The Gm17PK-YFP was predicted to localize to the cytoplasm. However, confocal imaging of the Gm17PK-YFP infiltrated leaves showed that it localizes to the plasma membrane (Figure 3.4 F). To confirm its localization, Gm17PK-YFP was co-expressed with plasma membrane targeted CFP (PM-CFP). The overlap of CFP signal from PM-CFP and YFP signal from Gm17PK-YFP confirmed its localization to the plasma membrane.

The subcellular localization study of Gm17PPD demonstrated its localization to the plastid as predicted. To confirm the localization, co-expression of Gm17PPD-YFP and plastid targeted CFP (PT-CFP) was performed. The overlap of signals from PT-CFP and Gm17PPD-YFP revealed that Gm17PPD localizes to the plastid (Figure 3.4 G).
**Figure 3.4 Subcellular localization of candidate GmPKs.** *N. benthamiana* leaves were co-transformed with candidate GmPKs fused with yellow fluorescent protein and organelle targeted signals fused with cyan fluorescent protein, followed by confocal microscopy. The signals from CFP and YFP are merged to conform the location of GmPKs. A) Gm02PK-YFP and ER-CFP B) Gm04PK-YFP and ER-CFP C) Gm08PK.1-YFP and NLS-CFP D) Gm08PK.2-YFP and NLS-CFP E) Gm14PK-YFP and ER-CFP F) Gm17PK-YFP and PM-CFP G) Gm17PPD-YFP and PT-CFP. Scale bars indicate 20 μm.
3.6 Candidate GmPKs interact with GmMYB176 in planta

To confirm protein-protein interaction between candidate GmPKs and GmMYB176, bimolecular fluorescence complementation (BiFC) assays were performed. As explained in section 2.5.1, protein-protein interaction allows the YN and YC domains to come into close proximity to each other, resulting in fluorescence. Importantly, the subcellular location of the interaction can also be determined using this method.

The protein-protein interaction was assayed using combinations such as GmPK-YN with GmMYB176-YC and GmMYB176-YC with GmPK-YN. The combination of GmPK-YN with GmMYB176-YC demonstrated that candidates Gm02PK, Gm04PK, Gm08PK.1, Gm08PK.2, Gm14PK and Gm17PPD interacted with GmMYB176 in the nucleus whereas Gm17PK interacted with GmMYB176 both in the nucleus and in the cytoplasm (Figure 3.5). However, in context of reciprocal combination, there was a discrepancy - only Gm08PK.1, Gm08PK.2, Gm17PK and Gm17PPD showed interaction with the GmMYB176 (Appendix D). The interactions were weak compared to the previous combinations.
Figure 3.5 GmMYB176 interact with candidate GmPKs in the nucleus. *A. tumefaciens* GV3101 containing translational fusions of a candidate GmPK with the N-terminal half (YN) and GmMYB176 with the C-terminal (YC) half of yellow fluorescent protein (YFP) were co-infiltrated into *N. benthamiana* leaves followed by confocal microscopy. Scale bars indicate 40 μm.
3.7 Recombinant Gm08PK.1 and Gm08PK.2 protein expression and purification

In order to obtain purified protein for kinase assay, candidate protein kinases were expressed in *E. coli*. Among seven candidate protein kinases, Gm08PK.1 and Gm08PK.2 were first selected for expression as they localize to the nucleus. GmMYB176-specific protein kinases are speculated to localize either in the cytoplasm or in the nucleus as GmMYB176 are found in these compartments (Li et al., 2012).

Protein expression can be affected by multiple factors such as temperature, IPTG concentration and induction time (Francis, 2010). To optimize the expression of the recombinant Gm08PK.1 (expected size of 106.4 kDa) and Gm08PK.2 (expected size of 91.7 kDa) protein, they were induced with 0.3 mM IPTG at 30ºC, 25ºC, and 18ºC for 6 h, 11 h, and 24 h, respectively. Induction with 0.3 mM IPTG at 18ºC for 24 h was optimal for the expression of recombinant Gm08PK.1 protein (Figure 3.6). Induction with 0.3 mM IPTG at 25ºC for 11 h was optimal for recombinant Gm08PK.2 expression (Figure 3.7). Therefore, these conditions were identified as optimal for induction and used for all future expression.

Both recombinant Gm08PK.1 and Gm08PK.2 proteins contain a 6X His tag both at the N- and C-terminus. His tags have an affinity for Ni²⁺; thus, Ni-affinity chromatography was used to purify the recombinant Gm08PK.1/Gm08PK.2 proteins. In this assay, the recombinant protein interacted reversibly with the Ni²⁺ attached to the matrix and could be easily eluted using competitive ligands such as imidazole. Elution buffer containing 50-500 mM imidazole was applied to purify the recombinant proteins. The recombinant protein Gm08PK.1 and Gm08PK.2 could be purified using 150-200 mM imidazole (Figure 3.8) and 200-500 mM imidazole concentrations (Figure 3.9), respectively. The overall yield of recombinant Gm08PK.1 was 8 mg/L and, of Gm08PK.2 was 3.5 mg/L.
Figure 3.6 Optimization of recombinant Gm08PK.1 protein expression in *E. coli* Rosetta(DE3). The bacteria containing pET32a(+)−Gm08PK.1 were induced using 0.3 mM IPTG at 18°C for 24 h, 25°C for 11 h and 30°C for 6 h. Total soluble protein (15 μL) was separated on a 8% SDS-PAGE and stained with Coomassie blue. Red arrow indicates Gm08PK.1 recombinant protein with a high yield in the soluble fraction. Expected recombinant Gm08PK.1 protein size is 106.4 kDa. UI: Uninduced; I: Induced.
Figure 3.7 Optimization of Gm08PK.2 recombinant protein expression in *E. coli* Rosetta(DE3). The bacteria containing pET32a(+)–Gm08PK.2 were induced using 0.3 mM IPTG at 18°C for 24 h, 25°C for 11 h and 30°C for 6 h. Total soluble protein (15 μL) was separated on a 8% SDS-PAGE and stained with Coomassie blue. Red arrow indicates Gm08PK.2 recombinant protein with high yield in the soluble fraction. Expected recombinant Gm08PK.2 protein size is 91.7 kDa. UI: Uninduced; I: Induced.
Figure 3.8 Purification of recombinant protein Gm08PK.1. *E.coli* Rosetta(DE3) containing pET32a(+) -Gm08PK.1 was induced using 0.3 mM IPTG at 18°C for 24 h. Crude soluble cell extracts (Lysate) containing recombinant Gm08PK.1 was passed through a Ni²⁺ charged Ni Sepharose High Performance column followed by washing. After washing, the recombinant protein was eluted with elution buffer containing different concentrations of imidazole a) 50-100 mM and b) 150-500 mM. Eluates (15 µL) was separated on a 8% SDS-PAGE and stained with Coomassie blue. Eluate from Lane 1-5 from 200 mM and Lane 1 from 500 mM were collected.
Figure 3.9 Purification of recombinant protein Gm08PK.2. *E.coli* Rosetta(DE3) containing pET32a(+) -Gm08PK.2 was induced with 0.3 mM IPTG at 25°C for 11 h. Crude soluble cell extracts containing recombinant Gm08PK.2 was passed through a Ni²⁺ charged Ni Sepharose High Performance column followed by washing. The recombinant protein was eluted with buffer containing different concentrations of imidazole a) 50-100 mM and b) 150-500 mM. Eluates (15 µL) was separated on a 8% SDS-PAGE and stained with Coomassie blue. Eluate from Lane 1-5 from 200 mM and Lane 1-4 from 500 mM were collected.
3.8 Tag removal from recombinant protein kinases

The pET32a(+) expression construct contains cleavable thioredoxin protein (Trx tag), S tag and His tag, N-terminal to the candidate GmPK. The Trx tag helps to solubilize and stabilize the recombinant protein (Costa et al., 2014). Both S tag and His tag can be used to detect the recombinant protein in western blot and can also be used to purify the recombinant protein (Raines et al., 2000; S. Graslund 2008). These tags in the pET32a(+) can be cleaved using enterokinase. Enterokinase is a serine protease that cleaves after Lys residue at the cleavage site, Asp-Asp-Asp-Asp-Asp-Lys.

To obtain candidate protein kinase in its native form, the tag proteins from recombinant protein Gm08PK.1/Gm08PK.2 were cleaved using enterokinase. As shown in Figure 3.10, the enterokinase not only cleaved the tagged proteins at the specific site but also cleaved the candidate protein kinases at non-specific sites. The result was consistent for cleavage reactions carried out at 18°C, 25°C and 37°C. Therefore, I proceeded to perform *in vitro* kinase assays with the fusion tags present in the candidate protein kinases.
Figure 3.10 Enterokinase cleaves recombinant proteins, Gm08PK.1 and Gm08PK.2, at multiple non-specific sites. Recombinant protein Gm08PK.1 or Gm08PK.2 was incubated with enterokinase for the times indicated at 37°C. The reaction was stopped by heating the reaction at 95°C for 10 min. 20 µL from each cleavage reaction was separated on a 8% SDS-polyacrylamide gel and stained with Coomassie blue. Expected size of Gm08PK.1 is 87.7 kDa and of Gm08PK.2 is 73 kDa. EK: Enterokinase.
3.9 *In vitro* kinase assay

To determine the GmMYB176-specific protein kinase, *in vitro* kinase assays were performed with purified recombinant protein, Gm08PK.1/Gm08PK.2 in conjunction with GmMYB176 peptide, SMRKSVSMNNLSQY. Here, I used Universal Kinase Activity Kit to perform *in vitro* kinase assays. In this assay, a kinase phosphorylates its receptor, resulting in phosphorylated receptor and ADP. The coupling phosphatase then acts upon ADP to release β-phosphate from the generated ADP. Free inorganic phosphate in acidic environment forms a complex with malachite green phosphate detection reagents and produces green color. The development of green color is proportional to the amount of inorganic phosphate in the reaction. The signals are quantified by reading the absorbance of the reaction at 620 nm (Figure 3.12).

For *in vitro* kinase assay, I included three reactions: positive control, negative control and kinase reaction. In the positive control, ADP is used instead of ATP. Due to the release of β-phosphate orchestrated by coupling phosphatase, CD39L2, high signals were expected. In the negative control, the candidate protein kinase was not included in the reaction. The signals were expected to be low, as the hydrolysis of the ATP by protein kinase was shunned. The low signals were due to the hydrolysis of ATP by coupling phosphatase. The coupling phosphatase, CD39L2, is highly active on ADP than on ATP (Wu, 2011). Furthermore, the hydrolysis of ATP by coupling phosphatase can be predicted from the enzyme kinetics and can be subtracted as background signal. The kinase reaction included candidate protein kinase, GmMYB176 peptide, ATP and coupling phosphatase. The reactions were carried out as explained in the section 2.9.

For each candidate protein kinase, the optical density for positive, negative and kinase reactions were measured and are shown in Table 3.2. The experiment was repeated three times to verify its reproducibility. Both candidate protein kinase Gm08PK.1 and Gm08PK.2 showed significant OD$_{620}$ differences between negative controls and kinase reactions (Figure 3.11 and 3.12). The OD$_{620}$ value from kinase reaction was 4.72 times and 2.45 times greater than the negative control for Gm08PK.1 and Gm08PK.2, respectively. This suggests that both Gm08PK.1 and Gm08PK.2 phosphorylate GmMYB176 peptide *in vitro*. 
Figure 3.11 Principle of Universal kinase assay. Kinase transfers λ-phosphate of ATP to a receptor resulting in phosphorylated receptor and ADP. The coupling phosphatase, CD39L2, then releases β-phosphate from ADP. The free inorganic phosphate group in acidic environment forms a complex with Malachite green detection reagent and produces green color. The color development is quantified by reading absorbance at 620 nm.
Table 3.2 Optical density obtained for *in vitro* kinase assays at 620 nm.

<table>
<thead>
<tr>
<th>Candidate Protein kinase</th>
<th>Type of reaction</th>
<th>OD\textsubscript{620} obtained from Biological Replicates</th>
<th>Mean</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Gm08PK.1</td>
<td>Positive Control</td>
<td>0.412</td>
<td>0.28</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>Negative Control</td>
<td>0.091</td>
<td>0.077</td>
<td>0.082</td>
</tr>
<tr>
<td></td>
<td>Kinase Reaction</td>
<td>0.462</td>
<td>0.287</td>
<td>0.428</td>
</tr>
<tr>
<td>Gm08PK.2</td>
<td>Positive Control</td>
<td>0.263</td>
<td>0.27</td>
<td>0.302</td>
</tr>
<tr>
<td></td>
<td>Negative Control</td>
<td>0.086</td>
<td>0.117</td>
<td>0.129</td>
</tr>
<tr>
<td></td>
<td>Kinase Reaction</td>
<td>0.24</td>
<td>0.294</td>
<td>0.285</td>
</tr>
</tbody>
</table>
Figure 3.12 Candidate protein kinase Gm08PK.1 phosphorylates GmMYB176 peptide \textit{in vitro}. The reactions contain components as indicated in a kinase assay buffer. The reactions were incubated at room temperature (RT) for 10 min. Malachite green detection reagents were added to the reactions, followed by 20 min incubation. The absorbance of the color was measured at 620 nm. The graphs show the average of three biological replicates. Error bars represent standard error of the mean.
Candidate protein kinase Gm08PK.2 phosphorylates GmMYB176 peptide in vitro. The reactions contain components as indicated in a kinase assay buffer. The reactions were incubated at room temperature (RT) for 10 min. Malachite green detection reagents were added to the reactions, followed by 20 min incubation. The absorbance of the color was measured at 620 nm. The graphs show the average of three biological replicates. Error bars represent standard error of the mean.
3.10 Calculation of free inorganic phosphate generated in the kinase reaction

To determine the total amount of phosphate generated in the kinase reaction, corrected OD$_{620}$ is correlated with the OD$_{620}$ generated by the known amount of phosphate in the solution. As shown in Figure 3.13, a standard curve was generated with OD$_{620}$ obtained for known concentrations of phosphate (0 pmol - 5000 pmol). The corrected OD$_{620}$ was obtained by subtracting the OD$_{620}$ of negative control from the OD$_{620}$ of kinase reaction. By correlating to the standard curve, the amount of phosphate transferred to GmMYB176 peptide, by the candidate protein kinase Gm08PK.1 and Gm08PK.2 was 1170 pmol and 1036 pmol, respectively. In other words, out of 5.41*10$^{15}$ number of GmMYB176 molecules present in a kinase reaction, 0.745*10$^{15}$ GmMYB176 molecules were phosphorylated by Gm08PK.1. Similarly, 0.623*10$^{15}$ GmMYB176 molecules were phosphorylated by Gm08PK.2.
Figure 3.14 A phosphate standard curve generated from known concentrations of inorganic phosphate in a kinase assay buffer. The slope of the line regression line, 3725.9 pmol/OD, represents the amount of phosphate corresponding to a unit of absorbance at 620 nm.
Chapter 4: Discussion

4.1 Critical residues in the candidate protein kinase catalytic domains

The protein kinase catalytic domain is folded into a two lobed structure. Various crystal structures of protein kinases have shown that the conservation of the catalytic core. It is assumed that during phosphorylation all protein kinases attain a conformation similar to cAPK-α (Bossemeyer, 1995). The cAMP-dependent protein kinase catalytic subunit, α form (cAPK-α) is regarded as the universal structural reference for the catalytic core of protein kinases as it provided the first crystal structure of any protein kinase (Knighton et al., 1991). The conservation of the two lobed structure of the catalytic domain as well as the mechanism of transfer of phosphoryl groups, is orchestrated by conserved sequences in the catalytic domain (Bossemeyer, 1995).

The catalytic domain starts from the glycine-rich loop (Engh et al., 1996). Gly50, Gly52 and Gly55 form a glycine-rich loop in subdomain I, and are highly conserved in protein kinases. Gly50, Gly52 and Gly55 are conserved 94%, 99% and 85% of the time among protein kinases in the Swiss-Prot (31.0) database, respectively (Hemmer et al., 1997). Two of the glycine residues, Gly52 and Gly55 are conserved in Gm02PK, Gm04PK, Gm08PK.1, Gm08PK.2, Gm14PK and Gm17PK. However, Gly50 is not conserved in Gm08PK.1 and Gm08PK.2. The conservation of glycine residues shows that they have an important role in kinase activity. They have been known to bind nucleotide. Adenosine is sequestered in-between the two lobes of the catalytic domain and the phosphates of ATP are bound by a glycine-loop. The function of glycines is further demonstrated by substitution of individual glycine residues (Hemmer et al., 1997). Alteration of each glycine residue resulted in a decrease in ATP affinity and a decrease in the rate of kinase activity. Substitution of Gly52 showed the greatest effect whereas Gly55 showed the least effect (Hemmer et al., 1997; Grant et al., 1998). The alteration of Gly50 showed 50% of the effect that was seen when Gly52 was replaced. This suggests that Gm08PK.1 and Gm08PK.2, which lack Gly50, may have reduced kinase activity.
The other conserved and the best characterized residue in the catalytic domain, lying in subdomain II, is Lys72. It is conserved in all aligned candidate GmPKs. This critical residue plays a role in the phosphoryl transfer reaction. The site directed mutation of Lys72 resulted in the loss of protein kinase activity (Hanks et al., 1988; Bossemeyer, 1993). In addition to its primary function to facilitate phosphoryl transfer, it may also contact α- and β-phosphoryl groups of the bound ATP (Bossemeyer, 1993; Adams, 2001). The interaction between Lys72 and the α- and β-phosphate of ATP is further stabilized by a salt bridge formed by Glu91 with Lys72 (Hanks and Hunter, 1995). The Glu91 residue, lying in subdomain III, is conserved in all aligned GmPKs.

Three conserved residues lie in subdomain VIB: Asp166, Lys168 and Asn171. These three residues are found conserved in Gm04PK, Gm08PK, Gm14PK and Gm17PK. The three conserved residue play a direct role in catalysis, therefore this domain is termed the catalytic loop (Smith et al., 1999). The catalytic loop begins with Asp166, which is believed to accept the proton from the hydroxyl group of the substrate amino acid during phosphoryl transfer (Hanks and Hunter, 1995). In order to confirm the function of Asp166, the residue was substituted with alanine (Gibbs and Zoller, 1991). Unlike Lys72, the mutation of Asp166 did not result in a loss of kinase activity. However, it did result in a decrease in both binding affinity for substrate and in catalytic activity (Gibbs and Zoller, 1991). The other residue in the catalytic loop, Lys168, contributes to catalysis by stabilizing the transition state of phosphoryl transfer (Madhusudan et al., 1994). It binds with the γ-phosphate of ATP both before and after phosphoryl transfer. The decrease in efficiency of catalysis by 50 fold after substitution of Lys168 with alanine demonstrates the importance of Lys168 in catalysis (Cheng et al., 2005). The third conserved residue of the catalytic loop, Asn171, forms hydrogen bond to the backbone carbonyl of Asp166 and chelates the secondary Mg$^{2+}$ ion that bridges α- and γ-phosphate of ATP (Zheng et al., 1993). The mutation of Asn171 severely inhibited kinase activity (Barylko et al., 2002).

The Asp184 residue, lying in subdomain VII, also chelates primary Mg$^{2+}$ (Adams, 2001). The Mg$^{2+}$ contributes to enhance the binding affinity of ATP and the formation of the transition state of the phosphoryl transfer reaction (Yu et al., 2011). The chelation of Mg$^{2+}$ by Asp184 either neutralizes the charge of γ-phosphate or orients γ-phosphate for transfer
to the hydroxyl acceptor. The site directed mutation of Asp184 result in a loss of protein kinase activity (Gibbs and Zoller, 1991). This critical residue is conserved in Gm04PK, Gm08PK, Gm14PK and Gm17PK, but not in Gm02PK.

The residues which are conserved in protein kinases, but whose functions are still unknown are Gly186, in subdomain VII; Glu208, in subdomain VIII; Asp220 and Gly225, in subdomain IX; and Arg280, in subdomain XI. All these residues are conserved in Gm04PK, Gm08PK.1, Gm14PK and Gm17PK. In Gm08PK.2, all the residues are conserved except Arg280. As mentioned earlier, the nearly invariant Arg280 residue’s function during catalysis is not known, but it is undoubtedly important as a replacement to alanine affects the kinase activity (Gibbs and Zoller, 1991). The lack of Arg280 in Gm08PK.2 could possible mean it has low kinase activity.

Among the aligned candidate protein kinases, Gm02PK lacks Asp166, Asn171, Asp184, Gly186, Glu208, Asp220, Gly225 and Arg280. All these residues are undoubtedly important for catalysis. Within these residues, Asp166, Lys168 and Asp184 are known to have a direct role in catalysis and the kinase activity is severely affected if the residue is replaced. Kinases that lack any of the critical residues - Lys72, Asp166, Asp184, are termed pseudokinases (Manning et al., 2002). Pseudokinases are predicted to be inactive (Zeqiraj and van Aalten, 2010). However, there are pseudokinases such as Ste20 related adaptor α isoform (STARDα) and Integrin-linked kinase (ILK) that lack Asp184, but are still showing kinase activity (Zeqiraj and van Aalten, 2010). There is a thus a possibility that pseudokinase Gm02PK could be inactive or may have a low kinase activity.

4.2 Candidate GmPKs family and their potential role

An examination of the catalytic domain of a protein kinase can reveal substrate specificity and the mode of substrate regulation. Kinase domain searches predicted that the candidate protein kinases have different protein kinase domains. For example: Gm02PK and Gm04PK are predicted to possess ethylene responsive protein kinase domain that includes CTR1. The name CTR1 comes from triple response – inhibition of elongation, radial swelling of the epicotyl, and altered response to gravity due to the exposure of ethylene to plants in the dark (Kieber, 1997). Ethylene plays many diverse roles in plants –
germination, senescence, abscission, flowering, fruit ripening and biotic and abiotic stress (Ju and Chang, 2012). The ethylene responsive protein kinase, CTR1, acts as a negative regulator in the ethylene signaling pathway by inhibiting downstream signaling (Huang et al., 2003).

Gm02PK is also a pseudokinase as it lacks critical residues in the catalytic domain. Until now, there has been no report of a pseudokinase with the ethylene responsive protein kinase CTR1 like domain. It could be possible that Gm02PK, similar to other pseudokinases such as KSR, may only act as a scaffold to bring together components of and regulate downstream signaling (Zeqiraj and van Aalten, 2010).

Gm08PK.1 and Gm08PK.2 are predicted to be cyclin-dependent protein kinases (CDK). Based upon the conserved motif, these CDK proteins have been divided into eight classes: CDKA to CDKG. The CDKG kinases have the PLTSLRE motif. Since both Gm08PK.1 and Gm08PK.2 contain this conserved motif, they are members of the G-type CDK (Tank and Thaker, 2011). CDKG have been reported to be essential for synapsis and recombination during male gametogenesis, and to regulate pre-mRNA splicing during pollen wall formation (Huang et al., 2013; Zheng et al., 2014). Furthermore, they can also act as negative regulators of salinity tolerance and early flowering (Żabicki et al., 2013; Ma et al., 2015).

Gm14PK is predicted to be part of the receptor like kinase (RLK) family. This family contains an extracellular region containing a cysteine-rich galacturonan-binding (GUB) domain, followed by a wall-associated receptor kinase (WAD) domain along with a Ser/Thr domain at the C-terminus. Members of the RLK family play role in plant growth and development, disease resistance, environmental stress and self-incompatibility (Vij et al., 2008; Lehti-Shiu et al., 2009). Different extracellular signals are detected by RLK family members and transduce the signals downstream (Lehti-Shiu et al., 2009).

Gm17PK is predicted to be a calcium dependent protein kinase (CPK). It contains a CPK domain along with two calcium binding domains called EF hands. CPKs play roles in abiotic stress (cold, salinity and drought), biotic stress (fungal and bacterial infection) and development responses (Schulz et al., 2013).
Gm17PPD is predicted to be pyruvate orthophosphate dikinase (PPD). It catalyzes the conversion of pyruvate to phosphoenolpyruvate, the primary CO$_2$ acceptor in the C4 cycle (Usami et al., 1995). In C3 plants like soybean, their function in photosynthesis has not been reported. Instead, these proteins are hypothesized to function in stress response (Hýskova and Ryslava, 2013).

4.3 Subcellular localization of candidate GmPKs

The analysis of protein subcellular localization can provide insights into protein function (Scott et al., 2005). GmMYB176-specific protein kinases must localize to the same compartment as GmMYB176, (i.e. in the nucleus or in the cytoplasm) in order to phosphorylate GmMYB176. The study of the subcellular localization of candidate GmPKs was carried out in order to narrow down the search for GmMYB176-specific protein kinases.

The subcellular localization study revealed both Gm02PK and Gm04PK localizes to the endoplasmic reticulum. However, Gm02PK and Gm04PK do not have an ER target signal peptide nor an ER retention signal (KDEL/HDEL). Similar to Gm02PK and Gm04PK, the ethylene-responsive protein kinase, CTR1 in A. thaliana and Solanum lycopersicon lacks both target and retention signal but is still found in ER (Binder and Schaller, 2015). The study by Goa et. al (Goa. et al, 2003) found that the localization of ethylene-responsive protein kinase, CTR1 was associated with interaction with ethylene receptors. Ethylene receptors are predominantly found in the ER, where ethylene binding domains are present in the ER and signal output domains are cytosolic. This signal output domain interacts with CTR1 in order to transmit the signal (Binder and Schaller, 2015). The interaction is critical for activation of the ethylene-responsive protein kinase, CTR1 and for blocking downstream signals (Zhong and Chang, 2012).

Gm08PK.1 and Gm08PK.2 localize to the nucleus. The localization to the nucleus makes Gm08PK.1 and Gm08PK.2 the most probable candidates to phosphorylate GmMYB176, as unphosphorylated GmMYB176 are retained in the nucleus. The search for a nuclear localizing signal revealed the peptide sequence, PERKRKFSPI, which is responsible for nuclear localization in both Gm08PK.1 and Gm08PK.2. Gm08PK.1 and Gm08PK.2 fall in
the cyclin-dependent kinase family. Interestingly, CDKs (A-type, B-type and C-type) have been reported to localize to the nucleus (Mews et al.; Mironov et al., 1999; Kitsios et al., 2008).

Gm14PK localizes to the ER but does not contain an ER retaining signals such as KDEL/HDEL. Gm14PK is predicted to be a receptor like kinase. Thus, I searched for RLKs that localize to the ER. Most of the RLKs, that have been characterized, localize to the plasma membrane, which is expected as they detect extracellular signals (Shiu and Bleecker, 2001; Zuo et al., 2015). RLKs have not been reported to localize to the ER. One of the factors that affect localization of receptor kinase is dimerization. In A. thaliana, the RLK -CLV2 and CRN, when expressed separately, are non-functional and are found in the ER (Bleckmann et al., 2010). But, when co-expressed, they become functional and are found in the plasma membrane. These CLV2 and CRN proteins interact with one another earlier in the endoplasmic reticulum via a transmembrane domain. So, it could be possible that Gm14PK interacts with another, yet unknown RLK, to form a functional receptor kinase.

The subcellular localization study showed that Gm17PK localizes to the plasma membrane. The search for signal peptides revealed the absence of any plasma membrane target signals in Gm17PK. Since Gm17PK is predicted to be a calcium dependent kinase, I searched for localization of characterized CPKs. CPKs are found to localize to multiple compartments of cells: the nucleus and the cytosol, the plasma-membrane, the peroxisome and the endoplasmic reticulum (Dammann et al., 2003). Furthermore, the N-terminal region, containing the signal peptide of CPKs, greatly varies in terms of sequence and length. However, myristoylation sites (removal of methionine residues and the addition of the 14 carbon saturated fatty acid, myristic acid to the N-terminal glycine residue) and palmitoylation sites (addition of fatty acid, palmitic acid, to cysteine residue) are conserved among CPKs. It was reported that myristoylation and palmitoylation are required for the protein to localize to the plasma membrane (Dammann et al., 2003). In OSCPK2, a plasma membrane localized CPK of Oryza sativa, when glycine of the myristoylation motif MGCGQGS, and cysteine of the palmitoylation site MGSCCS are substituted with alanine, the protein was not localized to the plasma membrane (Martin and Busconi, 2000). Since
Gm17PK contains both myristoylation (MGCGS) and palmitoylation sites (MGCGS), they likely help Gm17PK to localize to the plasma membrane.

Even though no signal peptide was found in Gm17PPD, it was localized to the plastid. Gm17PPD is predicted to be a pyruvate orthophosphate dikinase. These proteins are known to have a role in photosynthesis in C4 plants. However, Gm17PPD is from soybean which is a C3 plant. In C3 plants, PPDs are reported to localize either to the chloroplast or to the cytoplasm (Aoyagi and Bassham, 1985; Hocking and Anderson, 1986).

4.4 Candidate GmPKs interact with GmMYB176 in nucleus

The subcellular localization study of the candidate GmPKs demonstrated that they localize to different subcellular compartments. Gm02PK, Gm04PK and Gm14PK localize to the endoplasmic reticulum while Gm08PK.1 and Gm08PK.2 localize to the nucleus. The other candidate protein kinase Gm17PK and Gm17PPD localize to the plasma membrane and to the plastid, respectively. As GmMYB176 localizes to the cytoplasm and to the nucleus, it was expected that only nuclear localized Gm08PK.1 and Gm08PK.2 would interact with GmMYB176. However, all candidate GmPKs interacted with GmMYB176 in the nucleus, except Gm17PK which interacted both in the nucleus and in the cytoplasm.

It has been demonstrated previously that the phosphorylated fraction of GmMYB176 are retained in the cytoplasm by their interaction with 14-3-3 proteins. In contrast, unphosphorylated fraction of GmMYB176 is found in the nucleus (Li et al., 2012). 14-3-3 proteins are known to affect subcellular localization of client proteins (Muslin and Xing, 2000). Therefore, if the candidates GmPKs are phosphoproteins, and can interact with 14-3-3, it is possible that this interaction may alter their subcellular localization, thus providing an opportunity to interact with GmMYB176 in the nucleus for its phosphorylation. There are studies where phosphorylation affected the localization of protein kinases (Shah et al., 2001). In addition, protein kinases like cyclin-dependent protein kinase, calcium dependent protein kinase and ethylene response protein kinase, CTR1 are reported to interact with 14-3-3 proteins (Feng et al., 2005). Thus, 14-3-3 binding sites in candidate protein kinases were searched using online tool Scansite 3 (http://scansite3.mit.edu). The highest scoring predicted pST binding sites in the candidate GmPKs are listed in Table 4.1.
### Table 4.1 pST binding site for candidate GmPKs

<table>
<thead>
<tr>
<th>Candidate GmPKs</th>
<th>pST binding motif</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gm02PK</td>
<td>KSNYRQSTLSSSSQS</td>
</tr>
<tr>
<td>Gm04PK</td>
<td>QRFDRQVSLPRLSSG</td>
</tr>
<tr>
<td>Gm08PK.1</td>
<td>ERGARGRTSSGDDYP</td>
</tr>
<tr>
<td>Gm08PK.2</td>
<td>ERGARGRTSSGDDYP</td>
</tr>
<tr>
<td>Gm14PK</td>
<td>CFAKHHPSCPPLSSCG</td>
</tr>
<tr>
<td>Gm17PK</td>
<td>SKEKSKTEFTTSYG</td>
</tr>
<tr>
<td>Gm17PPD</td>
<td>VILVRNETSPEDVGG</td>
</tr>
</tbody>
</table>

Underline amino acid are potential phosphorylation site critical for 14-3-3 binding.
GmMYB176 interaction with several candidate protein kinases suggests that there could be more than one GmMYB176-specific protein kinase. Transcription factors are often phosphorylated at multiple sites (Komeili and O'Shea, 1999). Transcription factor such as ERK-1 is phosphorylated at two Ser and four Thr residues (Cruzalegui et al., 1999) and Pho4 is phosphorylated at five Ser residues (Komeili and O'Shea, 1999). Furthermore, motif scan analysis showed GmMYB176 has three potential phosphorylation sites - Ser29, Ser139 and Thr229 (Li et al., 2012). It could be possible that these three putative sites could be phosphorylated. Also, phosphorylation on each site could alter regulation of GmMYB176 differently. In the transcription factor, Pho4, the first and second phosphorylation events promote nuclear export whereas the third phosphorylation inhibits its nuclear import (Komeili and O'Shea, 1999). Thus, it could be possible that more than one candidate protein kinase may phosphorylate GmMYB176 in any one of putative serine residues, resulting into differential regulation of GmMBY176.

4.5 Non-specific cleavage by enterokinase

In order to perform the kinase assay with the native form of Gm08PK.1/Gm08PK.2, the tag proteins were cleaved using enterokinase. Enterokinase shows high specificity for cleaving proteins following Lys residue at its cleavage site, Asp-Asp-Asp-Asp-Lys. However, it cleaved Gm08PK.1/Gm08PK.2 at multiple non-specific sites. It has been reported that an enterokinase may cleave a protein after basic amino acid (Lys or Arg), that is preceded by an acidic amino acid (Asp or Glu) (Light and Janska, 1989; Shahravan et al., 2008).

Though the fusion tags in the recombinant protein may affect the enzymatic activity of the candidate protein kinase, there are numerous studies showing enzymatic assays performed along with the fusion tags (Bruneau et al., 2006; Panek et al., 2013; Bejger et al., 2014). I, therefore, proceeded to in vitro kinase assay along with the fusion tags.

4.6 Gm08PK.1 and Gm08PK.2 phosphorylate GmMYB176 peptide in vitro

GmMBY176 are predominantly localized to the nucleus and partially in the cytoplasm (Li et al., 2012). It was speculated that GmMYB176-specific protein kinases should localize
to these compartments in order to phosphorylate GmMYB176. The subcellular localization of candidate protein kinases demonstrated that both Gm0PK.1 and Gm08PK.2 localize to the nucleus. Furthermore, these candidate protein kinases interact with the GmMYB176 in the nucleus. These premises led to an assumption that Gm08PK.1 or/and Gm08PK.2 may phosphorylate GmMYB176. Thus, Gm08PK.1 and Gm08PK.2 were expressed in bacteria and purified. The purified recombinant protein Gm08PK.1/Gm08PK.2 in conjunction with GmMYB176 peptide was used to perform in vitro kinase assay.

The in vitro kinase assay, here performed, is based upon the measurement of optical density, which is proportional to the total amount of free phosphate generated in the kinase reaction (Wu, 2011). Both candidate protein kinases, Gm08PK.1 and Gm08PK.2 gave higher optical density values than the negative controls. This result suggests that these candidate protein kinases phosphorylate GmMYB176 peptide in vitro.

The GmMYB176 peptide, SMRKSVSMMNLSQY, can be phosphorylated at any of the four serine residues (underlined). From the obtained preliminary results, one could not conclude which residue was phosphorylated. As serine-29 phosphorylation in GmMYB176 is solely responsible for intracellular localization of GmMYB176, it is essential to identify serine residue that is phosphorylated in the kinase reactions. In order to figure out whether serine-29 is phosphorylated, in vitro kinase assay needs to be performed with GmMYB176 peptide where serine-29 is replaced with alanine.

In this in vitro kinase assay, values of OD_{620} are proportional to the amount of ATP hydrolyzed by candidate protein kinases. The efficiency to hydrolyze ATP by recombinant protein kinases may not represent to that of the native proteins. Two of factors may hinder the efficiency of recombinant proteins Gm08PK.1 and Gm08PK.2. First, the fusion tags present in the recombinant proteins Gm08PK.1 and Gm08PK.2 may affect their enzymatic activity. In a comparative study of trehalose synthase with His tags and wild-type trehalose, wild-type trehalose showed higher affinity to its substrate and higher efficiency for catalysis than trehalose with tags (Panek et al., 2013). Second, both Gm08PK.1/Gm08PK.2 are predicted to belong in the cyclin-dependent protein kinases that are generally activated by protein cyclin. Baculovirus-expressed cyclin A1/cdk2 (cyclin-dependent protein kinase)
complexes were able to phosphorylate human as well as murine B-myb in vitro (Müller-Tidow et al., 2001). In order to know the effect of cyclin on enzymatic activity of candidate protein kinases, in vitro kinase assay should be performed with and without cyclin.
Chapter 5: Perspective and future work

Isoflavonoids are important for human health and for stress resistance in plants. To increase the production of isoflavonoids in plants, an understanding of the biosynthetic pathway is critical. One of the key steps in the isoflavonoid pathway is the formation of chalcone by CHS8, which is regulated by GmMYB176 (Yi et al., 2010b). GmMYB176 interacts with 14-3-3 proteins, and serine-29 in GmMYB176 is critical for the interaction (Li et al., 2012). It is hypothesized that GmMYB176 is phosphorylated at serine-29 by a specific protein kinase. GmMYB176 interacting protein kinases were identified in a co-immunoprecipitation experiment and are being characterized.

From the in silico analysis, it was known that Gm02PK carries the least potential to be an active kinase, as it lacks three critical residues for phosphorylation. The remaining candidate protein kinases have critical residues conserved, and thus likely have the ability to phosphorylate its substrate. Furthermore, these seven candidate protein kinases show interaction with GmMYB176 in the nucleus in planta. The subcellular localization study revealed that Gm08PK.1 and Gm08PK.2 localize to the nucleus while Gm02PK, Gm04PK and Gm14PK localize to the endoplasmic reticulum. The other protein kinases Gm17PK and Gm17PPD localize to the plasma membrane and to the plastid, respectively. It is not known what molecular mechanism(s) is responsible for the observation that GmPKs localizing to different compartments of the cell still interact with GmMYB176 in the nucleus.

As Gm08PK.1 and Gm08PK.2 localize to the nucleus, and show interaction with the GmMYB176, they are selected to test if they are GmMYB176-specific protein kinase. Recombinant Gm08PK.1/Gm08PK.2 proteins were produced and purified. The purified recombinant protein, Gm08PK.1/Gm08PK.2 was used in conjunction with GmMYB176 peptide (SMRKSVMNNLSQY) in the in vitro kinase assay. Both of the candidate protein kinases showed comparatively high optical density than negative control in the in vitro kinase assays. This result suggests that both candidate protein kinases phosphorylate GmMYB176 in vitro. However, experiments with additional controls should be conducted to conclude Gm08PK.1 or Gm08PK.2 phosphorylates GmMYB176 in vitro. Furthermore, if phosphorylation in GmMYB176 is occurring in the kinase reactions, it is essential to
identify serine residue that is phosphorylated in the GmMYB176 peptide. It is possible that Gm08PK.1/Gm08PK.2 phosphorylate serine residue in GmMYB176 other than serine-29. In that case, in order to identify the GmMYB176-specific protein kinase that phosphorylates at serine-29, *in vitro* kinase assay should be performed with each remaining candidate protein kinases.
References


Appendices

Appendix A. Characteristics of candidate protein kinases

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Locus name</th>
<th>Coding sequence (nt)</th>
<th>Predicted protein molecular weight (kDa)</th>
<th>Predicted Subcellular localization</th>
<th>Predicted protein kinase classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gm02PK</td>
<td>Glyma.02G215300</td>
<td>2793</td>
<td>104.11</td>
<td>Nucleus</td>
<td>Ethylene response protein kinase CTR1</td>
</tr>
<tr>
<td>Gm04PK</td>
<td>Glyma.04G096000</td>
<td>2784</td>
<td>103.51</td>
<td>Nucleus</td>
<td>Ethylene response protein kinase CTR1</td>
</tr>
<tr>
<td>Gm08PK.1</td>
<td>Glyma.08G102600</td>
<td>2238</td>
<td>87.7</td>
<td>Nucleus</td>
<td>Cyclin-dependent protein kinase</td>
</tr>
<tr>
<td>Gm08PK.2</td>
<td>Glyma.08G102600</td>
<td>1962</td>
<td>73.05</td>
<td>Nucleus</td>
<td>Cyclin-dependent protein kinase</td>
</tr>
<tr>
<td>Gm14PK</td>
<td>Glyma.14G118800</td>
<td>2106</td>
<td>80.86</td>
<td>Plastid</td>
<td>Receptor like protein kinase</td>
</tr>
<tr>
<td>Gm17PK</td>
<td>Glyma.17G013800</td>
<td>1617</td>
<td>59.77</td>
<td>Cytoplasm</td>
<td>Calcium dependent protein kinase</td>
</tr>
<tr>
<td>Gm17PPD</td>
<td>Glyma.17G020600</td>
<td>2853</td>
<td>104.28</td>
<td>Plastid</td>
<td>Pyruvate phosphate dikinase</td>
</tr>
</tbody>
</table>

nt: nucleotide; kDa: kilo-Dalton.
Appendix B. Multiple sequence alignment of catalytic domains of candidate protein kinases

<table>
<thead>
<tr>
<th>Protein</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gm02 PK</td>
<td>669 PVDPLRKKRVRASQHSLGVLVLEYGSGKSLKSEKLRVQFQYDEQFQDGRKQKKEHHMKQGKMKK--</td>
</tr>
<tr>
<td>Gm04 PK</td>
<td>653 SWDNLRTKKEVRASQHSLGVLVLEYGSGKSLKSEKLRVQFQYDEQFQDGRKQKKEHHMKQGKMKK--</td>
</tr>
<tr>
<td>Gm14 PK</td>
<td>380 KIMTKFFKCSQHSLGVLVLEYGSGKSLKSEKLRVQFQYDEQFQDGRKQKKEHHMKQGKMKK--</td>
</tr>
<tr>
<td>Gm08 PK.1</td>
<td>397 -DBR-LKKEVRASQHSLGVLVLEYGSGKSLKSEKLRVQFQYDEQFQDGRKQKKEHHMKQGKMKK--</td>
</tr>
<tr>
<td>Gm08 PK.2</td>
<td>397 -DBR-LKKEVRASQHSLGVLVLEYGSGKSLKSEKLRVQFQYDEQFQDGRKQKKEHHMKQGKMKK--</td>
</tr>
<tr>
<td>Gm17 PK</td>
<td>88 -KYS-LKKEVRASQHSLGVLVLEYGSGKSLKSEKLRVQFQYDEQFQDGRKQKKEHHMKQGKMKK--</td>
</tr>
<tr>
<td>cAPK-α</td>
<td>42 -DBR-LKKEVRASQHSLGVLVLEYGSGKSLKSEKLRVQFQYDEQFQDGRKQKKEHHMKQGKMKK--</td>
</tr>
<tr>
<td>FVPK-1</td>
<td>227 -RHER-LKKEVRASQHSLGVLVLEYGSGKSLKSEKLRVQFQYDEQFQDGRKQKKEHHMKQGKMKK--</td>
</tr>
</tbody>
</table>

![Multiple sequence alignment of catalytic domains of candidate protein kinases](image-url)
Appendix C. Subcellular localization of Gm08PK.1 merged with the confocal image under white light.
Appendix D. Gm08PK.1, Gm08PK.2, Gm17PK and Gm17PPD showed interaction with GmMYB176 in reciprocal combination.
Curriculum Vitae

Pravesh Lama

EDUCATION

2014 – Present  M.Sc candidate  Western University Biology
                London, Ontario, Canada

2006 – 2010  B.Tech. in Biotechnology  Kathmandu University
                Dhulikhel, Kavre, Nepal

AWARDS AND SCHOLARSHIPS

2014 – Present  Western Graduate Research Scholarship

WORK EXPERIENCE

2016 – Present  Graduate Research Assistant,
                Western University

2014 – 2015  Graduate Teaching Assistant
                Western University

CONFERENCES
