Characterization of SPX exclusive family members in plant Pi sensing and regulation

Wanjun Qi

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

Faculty of Biological Sciences

September 2016

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The work in Chapter 1 of the thesis has appeared in publication as follows:

1. Alison Baker, Stanislaus A. Ceasar, Antony J. Palmer, Jaimie B. Paterson, Wanjun Qi, Stephen P. Muench and Stephen A. Baldwin. Replace, reuse, recycle: improving the sustainable use of phosphorus by plants. (2015) *Journal of experimental botany*, **66**, 3523-3540.
2. Wanjun Qi, Stephen A. Baldwin, Stephen P. Muench and Alison Baker. Pi sensing and signalling: from prokaryotic to eukaryotic cells. (2016) *Biochemical Society Transactions*, **44**, 766-773.
3. Wanjun Qi, Stephen A Baldwin, Stephen P Muench and Alison Baker. Phosphate sensing and transport in plants: Investigation of SPX domain-containing proteins. *Aspects of Applied Biology* **124**, *Breeding Plants to Cope with Future Climate Change*, pp. 97-106.

Wanjun Qi was responsible for section 'PHO1 and its homologues' in the first publication and created Figure 2 and Table 2, the other authors contributed to the rest of this publication. Wanjun Qi wrote the majority of the second and third publications and created the figures, Stephen A Baldwin, Stephen P Muench and Alison Baker contributed to the text and corrections to figures.

Acknowledgments

Abstract

The macronutrient phosphorus is vital to plant growth but the readily absorbed form of inorganic phosphate (Pi) is often environmentally limited. To increase Pi acquisition and use efficiency, plants have developed strategies involving a group of proteins, characterized by a highly conserved N-terminal SPX domain, which play critical roles. In this project, 14 SPX domain-containing proteins were identified in potato (*Solanum tuberosum*) and classified into four plant SPX domain-containing protein families. Gene expression analysis of SPX-exclusive family members using quantitative real-time PCR showed *StSPX2*, *StSPX3* and *StSPX5* are transcriptionally up-regulated during Pi starvation responses, suggesting the pervasive involvement of SPX proteins in Pi regulation among different plant species. The P1BS cis-element was also found in the promoter region of potato SPX genes, suggesting these genes are likely to be regulated by a potato homologue of AtPHR1, a Myb-CC family transcription factor which binds this sequence in other plant species.

Recent studies in *Arabidopsis* demonstrated the SPX exclusive family protein AtSPX1 can compete with DNA to physically interact with the transcription factor AtPHR1 in a Pi-dependent way. However, the interaction mechanism is poorly understood. To investigate this, Glutathione-S-transferase-SPX1 and Maltose-binding-protein-PHR1 fusion proteins were expressed and purified from *E. coli*. AtPHR1 was shown to be a monomeric protein in solution that dimerizes upon associating with P1BS. SPR data also showed that AtSPX1 cannot displace AtPHR1 associated with DNA, but only interact with monomeric AtPHR1 in the presence of either 5 mM Pi or 500 μM InsP6. Therefore I propose a new model that in the Pi restored condition, AtSPX1 can bind to monomeric AtPHR1 and therefore regulate PSI gene expression by tuning the AtPHR1-DNA binding equilibrium. This regulation also generates a negative feedback loop on the expression of *AtSPX1* itself, providing a tight control of PSI gene expression under Pi sufficient conditions.

Table of Contents

List of Figures

List of Tables

Abbreviation - General

AC:	Affinity chromatography
AM:	Arbuscular mycorrhizal
Amp:	Ampicillin
APase:	Acid phosphatase
APS:	Ammonium persulphate
AUC:	Analytical ultracentrifuge
BCA:	Bicinchoninic acid
BSA:	Bovine serum albumin
CAI:	Codon adaptation index
CD:	Circular dichroism
CDK:	Cycline dependent kinase
CKI:	Cycline dependent kinase inhibitor
CPMV:	Cowpea mosaic virus
CR:	Cluster root
DMSO:	Dimethyl sulfoxide
DTT:	Dithiothreitol
EDTA:	Ethylenediaminetetraacetic acid
EMSA:	Electrophoresis mobility shift assay
EXS:	ERRD 1 – XPR1 – SYG1
GFP:	Green fluorescent protein
GndCl:	Guanidine hydrochloride
GST:	Glutathione-S-Transferase
HEPES:	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IEX:	Ion exchange chromatography
IMAC:	Immobilized metal affinity chromatography
InsP6:	Inositol hexakisphosphate
IP7:	Inositol heptakisphosphate
IPTG:	Isopropyl-β-D-thiogalactopyranoside
ITC:	Isothermal titration calorimetry

Kan:	Kanamycin
KSC:	Lysine surface cluster
LB:	Lysogeny broth
MBP:	Maltose-binding protein
MCA:	Micro-cuvette array
MFS:	Major facilitator superfamily
miRNA:	microRNA
MST:	Microscale thermophoresis
NAT:	Natural antisense transcript
NLA:	Nitrogen limitation adaptation
ORF:	Open reading frame
P:	Phosphorus
P1BS:	PHR1-binding sequences (PHR1-binding site)
PBC:	Phosphate binding cluster
PBS:	Phosphate-buffered saline
PCR:	Polymerase chain reaction
PHR:	Phosphate starvation response
Pi:	Inorganic phosphate
Pit	Phosphate inorganic transporter
PKA:	Protein kinase A
PMSF:	Phenylmethylsulfonyl fluoride
POCI:	Potato oligo chip initiative
PSI:	Phosphate starvation induced
PST:	Phosphate-specific transport
PUE:	Phosphate use efficiency
Q-PCR:	Quantitative polymerase chain reaction
RING:	Really interesting new gene
RT-PCR:	Reverse transcription polymerase chain reaction
SA:	Streptavidin
SB:	Super broth
SBP:	Streptavidin binding peptide
SDS-PAGE:	Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SEC:	Size exclusive chromatography
SLS:	Static light scattering
SPR:	Surface plasmon resonance
SPX:	Syg1 (Suppressor of yeast gpa1) – Pho81 (Phosphatase 81) – Xpr1 (Xenotropic and polytropic retrovirus receptor 1)
TAE:	Tris-acetate-EDTA
TBST:	Tris-buffered saline tween
TCEP:	Tris (2-carboxyethyl) phosphine
TE:	Tris-EDTA
TEMED:	Tetramethylethylenediamine
TF:	Transcription factor
VPT:	Vacuolar phosphate transporter
Vtc:	Vacuolar transporter chaperone
wHTH:	winged-helix-turn-helix

Abbreviation - Organisms

At: *Arabidopsis thaliana*

Ct: *Chaetomium thermophilum*

E. coli: *Escherichia coli*

Hs: *Homo sapiens*

Nb: *Nicotiana benthamiana*

Os: *Oryza sativa*

Sc: *Saccharomyces cerevisiae*

St: *Solanum tuberosum*

Chapter 1 Introduction

1.1 Phosphate (Pi) importance and its environmental availability

Phosphorus (P) is one of the most commonly found elements in the earth's crust and plays a key role in multiple biological processes. It is not only an essential structural component in nucleic acids and phospholipids, which serve as core constituents of living organisms, but it is also an indispensable factor in energy metabolism and signal transduction cascades that modulate gene expression patterns and physiological responses. Therefore, the availability of this essential nutrient is a growth limiting factor and can determine the yields of crop plants.

A 'phosphorus paradox' with regard to this critical plant nutrient has been highlighted for several decades (Bieleski, 1973; Marschner, 2012). In spite of the large reserve of total P in soils, the amount of its bioavailable form, inorganic phosphate (Pi), is frequently low. From 20% up to 80% of P in soils is found in organic molecules which are not directly utilizable by plants (Schachtman *et al.*, 1998). Of the bioavailable form Pi, the adsorption of phosphate onto the surface of clay particles and binding to aluminium and other metal ions also limits the average concentration of bioavailable Pi in the soil solution typically to around 1 µM and very rarely above 10 µM. This is in contrast to the normal Pi content in plants, which ranges from 5 mM to 20 mM (Bieleski, 1973).

The application of P fertilizers can compensate for the low endogenous Pi availability in soils and therefore increase soil fertility and crop yields. However, long-term phosphate fertilizer application creates the potential environmental risks of P-overloading in soils and its subsequent run-off that leads to eutrophication in watercourses (Lougheed, 2011). Moreover, the accumulation in croplands of potentially hazardous trace elements such as arsenic, cadmium, and lead (which can be present as contaminants in Pi fertilizers derived from rock phosphate) results in further environmental damage (Jiao *et al.*, 2012). In addition, 90 percent of agricultural P fertilizer comes from non-renewable mined rock phosphate. Readily accessible global reserves are being rapidly depleted, such that production is predicted to reach a peak by 2033 before gradually declining (Elser and Bennett, 2011; Cordell *et al.*, 2009) (Figure 1.1).
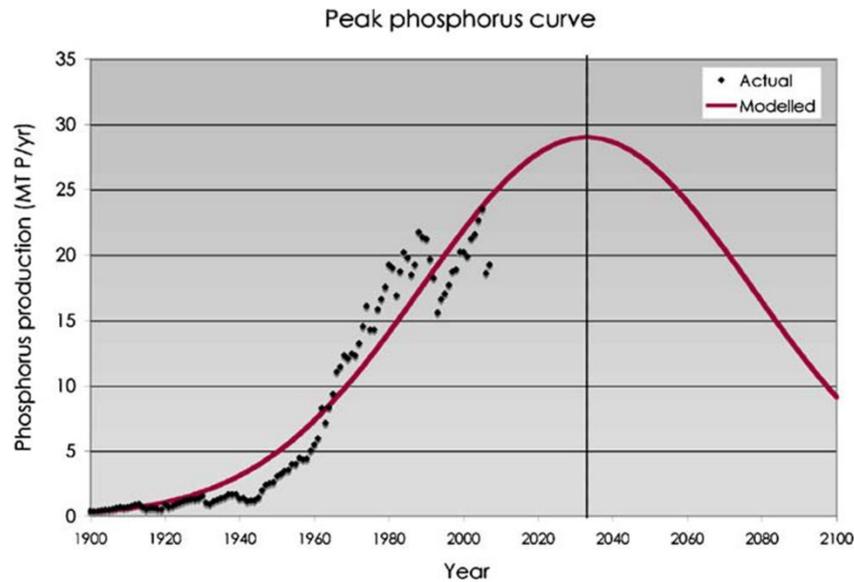
Figure 1.1 Prediction of peak global phosphorus production (Cordell *et al.*, 2009).

Analysis based on available industry data predicts the phosphorus production world-wild could peak by year 2033 (Cordell *et al.*, 2009).

The uncertainty of future supply, cost, and the environmental impact of run-off all indicate that an agricultural future relying on continued high level use of this inorganic fertilizer is neither environmentally nor economically sustainable. Therefore, a better understanding of how phosphate is sensed, acquired and utilized by living organisms, especially by plants is urgently required for the development of crop plants with greater phosphate use efficiency, which is of vital importance for more sustainable P management in agriculture.

In this chapter, the Pi sensing and signalling pathways are first discussed in the better understood unicellular bacteria and yeast models. It then moves on to discuss our current understanding of their plant counterparts. The key roles of SPX domain-containing proteins in Pi regulation is also highlighted together with some recent advances in the functional and structural studies on the SPX protein domain.

1.2 Pi regulation in *Escherichia coli*

In bacterial cells, Pi regulation typically involves Pi sensing and transport by a cytoplasmic membrane anchored protein complex as well as a two-component intracellular Pi signalling pathway, which modulates the expression of various genes that comprise the Pho regulon (Hsieh and Wanner, 2010) (Figure 1.2).

In *E. coli*, sensing and transport of environmental Pi is primarily accomplished by a high-affinity Phosphate-specific ABC transport (Pst) system, which consists of a periplasmic Pi binding protein (PstS), a Pi channel formed by 2 membrane proteins (PstA and PstC), and a cytoplasmic dimeric ATPase (PstB) which provides energy for Pi transport across the membrane (Chan and Torriani, 1996). Under conditions where environmental Pi availability is above 4 µM, Pi binding to PstS inhibits the inner membrane Pi sensory histidine kinase PhoR in a Pi-uptake independent manner (Hsieh and Wanner, 2010; Wanner, 1996), subsequently abolishing the activation of downstream Pho-controlled gene expression. Although the exact mechanism of this inhibition is unclear, it is well established that a chaperone/Hsp70-like protein PhoU plays a critical role. It has been proposed that PhoU regulates the activity of PhoR by forming a complex between the PstB protein of the Pi transporter and the PAS (Per-ARNT-Sim) domain of PhoR (Wanner, 1996; Oganesyan *et al.*, 2005). This hypothesis is also supported by a study from the Gardner group, which shows the binding of manganese and magnesium to dimeric PhoU via highly conserved residues which could be important for its membrane complex association (Gardner *et al.*, 2014). However, a more recent study in *Caulobacter crescentus* suggested rather than negatively regulating the Pho regulon, PhoU modulates Pi uptake by the Pst transporter (Figure 1.2) and prevents high intracellular Pi toxicity (Lubin *et al.*, 2016). Further evidence is needed to confirm the precise function of PhoU and whether it plays diverse roles in different organisms.
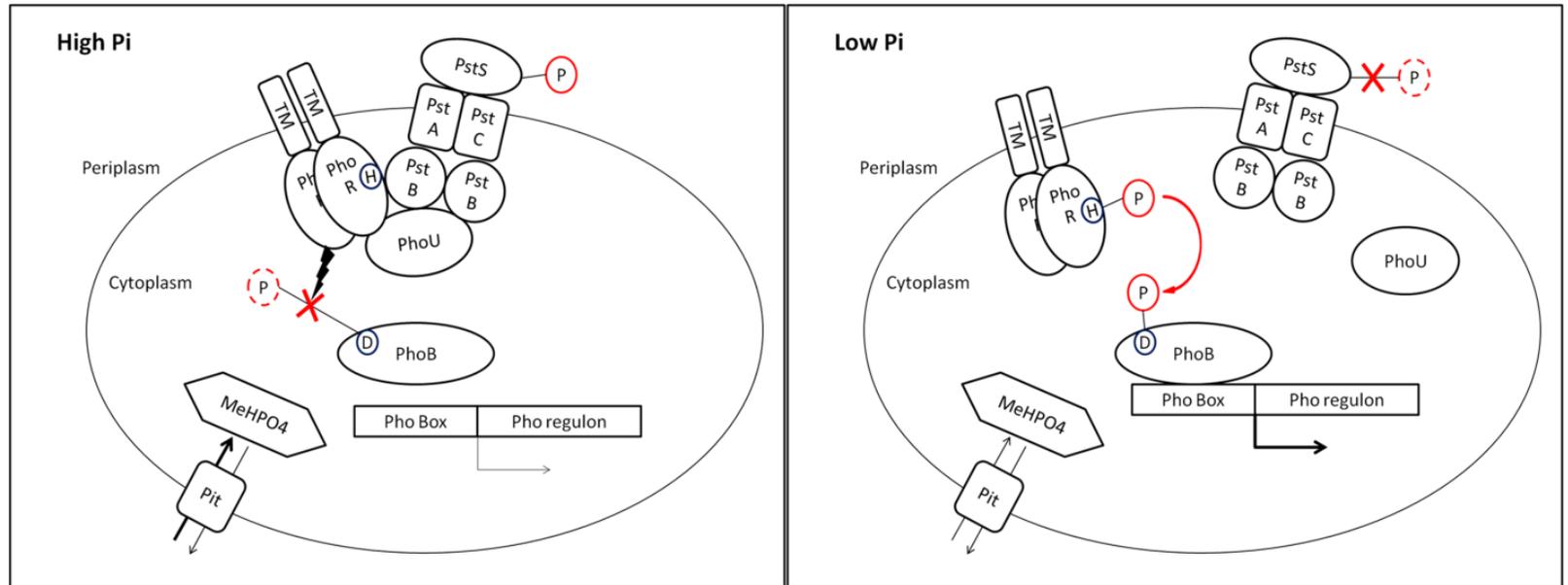
Figure 1.2 Schematic diagram of Pi regulation pathways in *E. coli* (Qi *et al.*, 2016).

Under high Pi conditions, binding of Pi to PstS results in inhibition of histidine kinase PhoR via the formation of a PhoR-PhoU-PstB protein complex, resulting in dephosphorylation and deactivation of the transcription factor PhoB. The low-affinity Phosphate inorganic transporter (Pit) system also mediates metal-phosphate complexes (MeHPO4) transport across the plasma membrane. Under low Pi conditions, PhoB is activated by phosphorylation on an aspartate residue (D) by the auto-phosphorylated histidine kinase PhoR and binds to specific DNA targets through the Pho box thereby allowing the activation of downstream Pho regulon genes.

Under Pi limited conditions, the repression on PhoR is released by disassociation from PhoU, thus allowing the auto-phosphorylation of this histidine kinase, and activation of its downstream response regulator PhoB (Hsieh and Wanner, 2010) (Figure 1.2). The transcription factor PhoB belongs to the OmpR/PhoB family, and is characterized by an N-terminal receiver domain and a C-terminal DNA binding domain which contains a unique wHTH (winged-Helix-Turn-Helix) domain (Nguyen *et al.*, 2015). A highly conserved Asp residue of PhoB is phosphorylated during low Pi stress and the phosphorylated PhoB subsequently activates the Pho regulon by binding to a specific Pho box sequence in the promoter region (Wanner, 1996; Nguyen *et al.*, 2015). Computational predictions based on known Pho box sequences revealed 96 putative Pho regulated genes in the genome of *Sinorhizobium meliloti* and more than 100 in *E. coli*. Among these sequences are the genes which encode the Pst Pi sensing/transporting complex, integral membrane protein PhoR and transcription factor PhoB itself (Yuan *et al.*, 2006), indicating the existence of a positive feedback loop of bacterial Pi regulation. The low-affinity Phosphate inorganic transporter (Pit) system also contributes to Pi uptake under Pi sufficient conditions (> 4 μM) (Figure 1.2). However, this system primarily transports metal-phosphate complexes and its expression is largely controlled by the external Zn^{2+} level (Jackson *et al.*, 2008). A recent study also shows the Pit system plays essential roles in *E. coli* copper tolerance by mediating Cu^{2+}-phosphate efflux from the cell (Grillo-Puertas *et al.*, 2014).

1.3   Pi regulation in *Saccharomyces cerevisiae*

As a model organism for eukaryotes, Pi regulation in yeast *Saccharomyces cerevisiae* (*S. cerevisiae*) has been intensively studied (Secco *et al.*, 2012b). In this system, external Pi availability is sensed by plasma membrane localized proteins Pho84 and Pho87. Pho84 is both a phosphate transporter and a phosphate sensor which can couple growth to Pi availability via activation of the protein kinase A pathway. Thus it is often termed a 'transceptor'. This dual function of sensing and transport is shared with other nutrient transporters in *S. cerevisiae* (Holsbeeks *et al.*, 2004). The subsequent activation of the Pi transporter system and maintenance of a stable intracellular Pi level is mediated by the phosphate-responsive signalling (PHO) pathway. This pathway is mainly regulated by Pho4, the transcription factor (TF); Pho80-Pho85, the cyclin dependent kinase (CDK) complex and Pho81, a cyclin dependent kinase inhibitor (CKI) (Figure 1.3).

Under Pi sufficient conditions, the nuclear localized Pho4 is phosphorylated by the Pho80-Pho85 CDK complex at multiple serine residues, resulting in disassociation from its cooperative TF Pho2 and its export from the nucleus into the cytosol (Komeili and O'Shea, 1999). When Pi is depleted, the conformation of Pho81, which constitutively associates with the Pho80-Pho85 complex, is changed through interaction with inositol heptakisphosphate (IP7) (Lee *et al.*, 2008) and therefore prevents the phosphorylation of Pho4 by Pho80-Pho85. The accumulated nuclear Pho4 interacts with Pho2 and activates the downstream Pi starvation induced (PSI) genes. A yeast genome wide analysis has revealed 22 genes that are transcriptionally regulated by this pathway, including those encoding high-affinity Pi transporters; Spl2, a negative regulator of low-affinity Pi transporters; secreted phosphatases; the vacuolar poly Phosphate (poly P) synthesis complex and Pho81 itself (Ogawa *et al.*, 2000).
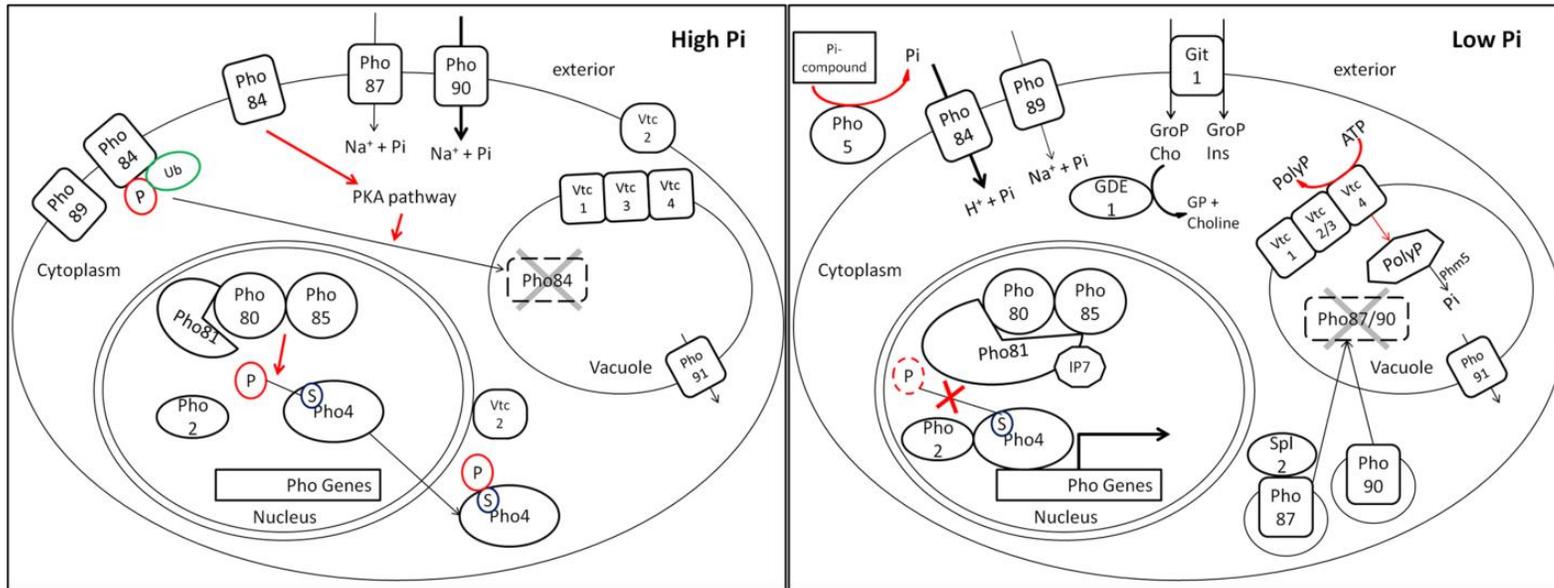
Figure 1.3 Schematic diagram of Pi regulation pathways in *S. cerevisiae* (Qi *et al.*, 2016).

Under high Pi conditions, Pho4 is phosphorylated on multiple serine residues (S) by the Pho80-Pho85 CDK complex and exported from the nucleus into the cytosol. The low-affinity Pi transporters Pho87 and Pho89 are responsible for Pi uptake and high-affinity Pi transporter Pho 84 is internalized and degraded. Under low Pi conditions, interaction with IP7 changes the conformation of Pho81 and prevents the phosphorylation of Pho4 by Pho80-Pho85. Accumulated nuclear Pho4 in concert with Pho2 activates the downstream Pi starvation induced (PSI) genes and leads to the upregulation of high-affinity Pi transporters as well as the internalization and degradation of low-affinity Pi transporters. Low Pi also triggers the formation of Vtc complex to increase the vacuolar accumulation of inorganic polyphosphate (polyP), as well as polyP hydrolysis by vacuolar polyphosphatase Phm5 and Pi transport from vacuole to cytosol by Pho91. Pi scavenging from macromolecules is also activated during Pi starvation, resulting in the import of glycerophosphoinositol (GroPIns) and glycerophosphocholine (GroPCho) through Git1 and the hydrolysis of GroPCho by Gde1.

The yeast Pi transporter system consists of two high-affinity Pi transporters (Pho84, Pho89) and two low-affinity Pi transporters (Pho87, Pho90) that are localized in the plasma membrane, as well as a vacuolar membrane localized Pho91 (Hürlimann *et al.*, 2007) (Figure 1.3). While the expression of Pho84 and Pho89 is highly induced by exposure to low Pi conditions, the low-affinity transporters are constitutively transcribed and are endocytosed for vacuolar degradation upon Pi depletion (Ghillebert *et al.*, 2011). Upon Pi addition to the medium, the Pi transceptors Pho84 and Pho87 sense high environmental Pi and transport it across the plasma membrane (Giots *et al.*, 2003). Pi-induced activation of the protein kinase A (PKA) pathway in turn leads to the phosphorylation and ubiquitination of Pho84 itself, and eventually results in its internalization and degradation in the vacuole (Lundh *et al.*, 2009) (Figure 1.3).

During Pi uptake, the intracellular Pi content is also balanced by the inorganic polyphosphate (poly P) synthesis and vacuolar accumulation, which is accomplished by four Vacuolar Transporter Chaperone (Vtc) family members (Ogawa *et al.*, 2000) (Figure 1.3). The structural and functional studies of Vtc4 have shown it to be a polyP synthesizing enzyme (Hothorn *et al.*, 2009) which catalyses polyP synthesis in the cytosol from ATP prior to its transport into the vacuole, whereas the fully functional Vtc complex requires the integrity of a heterotrimer Vtc1/2/4 or Vtc1/3/4. No Vtc member has been identified in plants or mammals, suggesting Pi storing in higher eukaryotic cells maybe by a different route. Pi starvation signals also trigger the scavenging of macromolecules for Pi remobilization. This procedure involves expression upregulation of secreted acid phosphatases (Pho5, Pho11, Pho12), vacuolar polyphosphatase (Phm5), glycerol phosphatase (Hor2) as well as the glycerophosphoinositol (GroPIns) / glycerophosphocholine (GroPCho) transporter (Git1) and glycerophosphodiester phosphodiesterase (Gde1), activity of which is also seen in bacterial, plant and mammalian cells (Ogawa *et al.*, 2000; Fisher *et al.*, 2005; Ohshima *et al.*, 2008; Van Der Rest *et al.*, 2004; Sok, 1998).

1.4 Plants Pi homeostasis maintenance and key protein family members

The nearly 1000-fold Pi concentration difference between soil solution (1-10 μM) and plant intracellular compartments (5-20 mM) confronts plants with persistent Pi paucity in the soil (Bieleski, 1973). Unlike bacteria and yeast, Pi homeostasis maintenance in plants requires different intracellular regulatory pathways in cells from varying tissue types, as well as intricate intercellular signalling and communication (Figure 1.4; Table 1.1). The core responses comprise enhanced acquisition of Pi from the soil environment as well as conservation and transporting of the element throughout the plant corpus (Figure 1.5) (Lin *et al.*, 2009; López-Arredondo *et al.*, 2014).

1.4.1 Enhanced Pi acquisition

Due to the immobility of plants, a variety of morphological changes are required to maintain their regular growth, development and metabolic activities under Pi limiting conditions, including alterations in the architecture of the root system (Sato and Miura, 2011). The changes in root architecture generally involve reduction in the growth of the primary root, increasing growth of secondary roots and increasing root hair length and density (Yuan and Liu, 2008), although diverse responses among different *Arabidopsis* accessions have been identified (Chevalier *et al.*, 2003).

The reduced growth capacity of the primary root, the first observed morphological change in the root architecture, results from a reduction in cell elongation and meristem activity (Sánchez-Calderón *et al.*, 2005). In *Arabidopsis*, the ER localized Low Phosphate Root 1 (LPR1, a multi copper oxidase) and Phosphate Deficiency Response 2 (PDR2, a P5 type ATPase) have been shown to play key roles in rhizosphere Pi sensing (Figure 1.4; Table 1.1). Quantitative trait loci analysis has demonstrated that the low Pi induced primary root cell length responses are under the dominant control of LPR1, although LPR1 is generally not involved in other classical Pi starvation responses (Reymond *et al.*, 2006). PDR2 has been shown to regulate stem cell differentiation and meristem activity through the expression modulation of transcription factor SCARECROW (SCR) (Ticconi *et al.*, 2009).
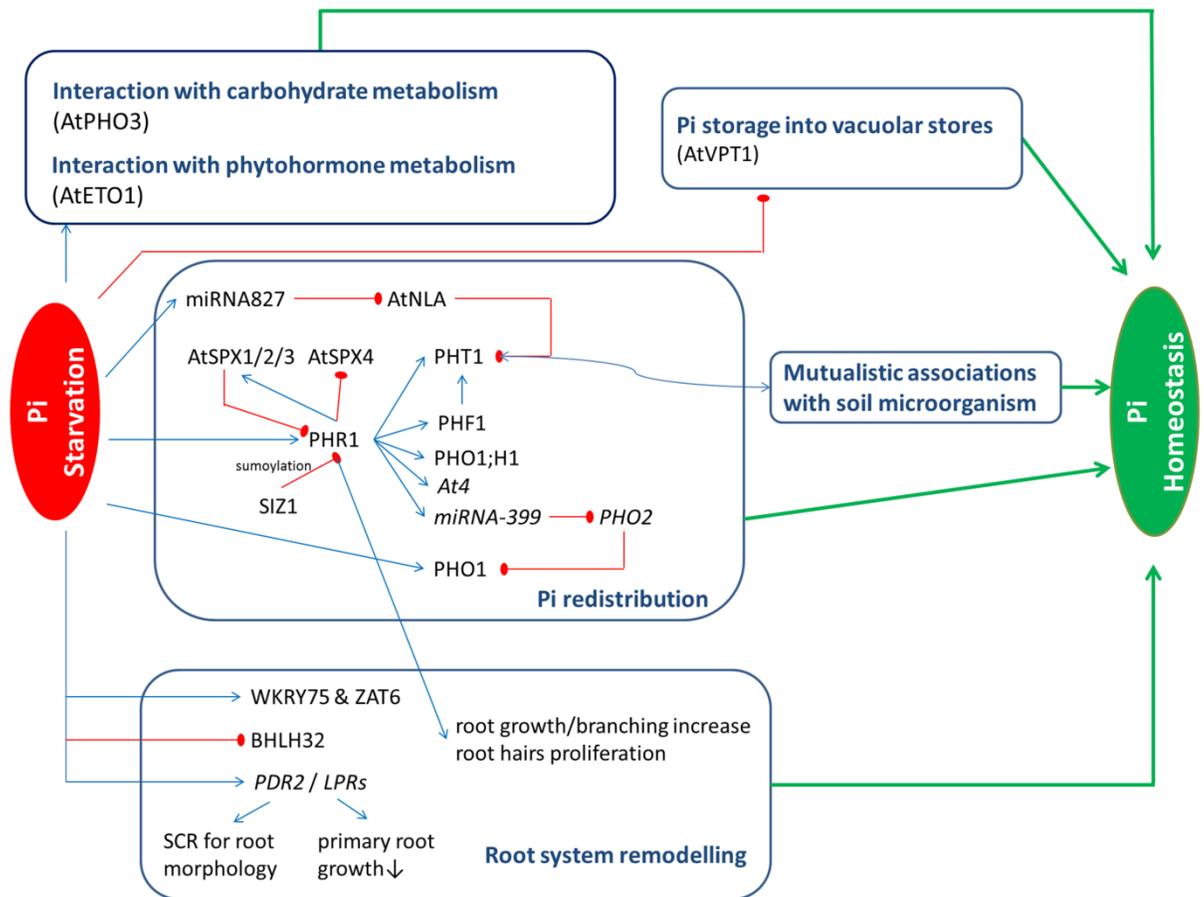
Figure 1.4 Regulation and control of key genes in the model plant *Arabidopsis* during Phosphate (Pi) starvation responses (Baker *et al.*, 2015).

In the presence of lowered environmental Pi concentrations, Root system remodeling is introduced, followed by the up-regulation of high affinity Pi transport systems (PHT1s) to increase Pi uptake from the soil, while specialized transporters (AtPHO1, AtPHO1;H1) are induced for the movement of Pi within the plant. A rigorous regulation system consisting of key transcription factor PHR1, post transcriptional regulation by non-coding RNAs and post translational regulation by protein trafficking and degradation is also involved for the functional integration of such transporters in response to Pi starvation. Blue arrowheads and red blunt-ends show positive and negative regulation, respectively.
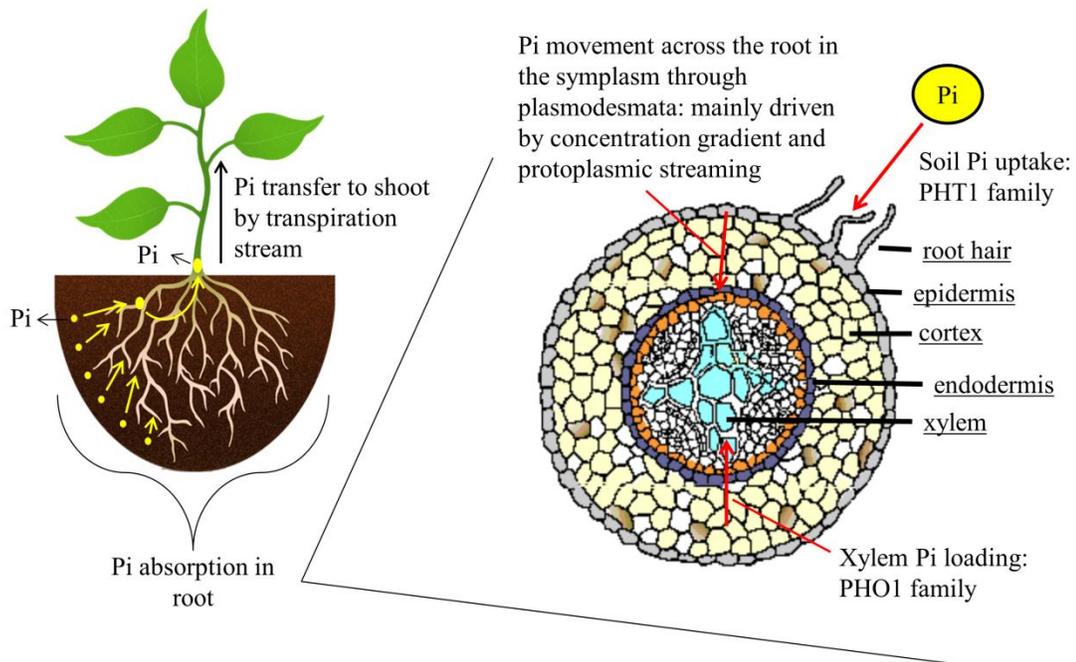
Figure 1.5 Outline map of Pi absorption and transport in plants

Bioavailable form of Pi is absorbed by positive uptake from the soil through PHT1 family members. Upon entering the epidermis, Pi travels across the cortex cells in the symplasm under the driving force from protoplasmic streaming and concentration gradient. Pi is loaded from the cortex into the xylem by PHO1 family members before being transferred to shoot tissues under the driving force from transpiration streaming.

General root architecture modulation in response to Pi starvation is also under the control of auxin regulation in a SIZ1 (SAP and Miz1, a SUMO-E3 ligase) dependent manner (Miura *et al.*, 2011). SIZ1 sumoylates the key transcription factor PHR1 (Figure 1.4). Mutants of SIZ1 exhibit more exaggerated prototypical root architecture changes, not only in the interrupted primary root growth, but also in the extensive development of secondary roots and root hairs. These exaggerated responses in *siz1* mutants can be restrained by treatment with auxin transport inhibitors, consistent with a negative regulatory role of SIZ1 in root architecture remodeling in response to the low Pi. Notably, the impaired primary root growth upon Pi deprivation can be restored by the supply of either Pi or its analogue phosphite, despite the lack of metabolic activity for the latter (Ticconi *et al.*, 2004), indicating that Pi sensing rather than Pi metabolism is locally controlling the root morphology reprogramming. A minority of plants, mainly from the Proteaceae family, also produces clustered lateral roots (cluster roots, CRs) to increase the nutrient absorbing area as well as to secrete organic acids and acid phosphatase, in order to survive on the most Pi impoverished soils (Müller *et al.*, 2015).

In addition to the root architecture reprogramming, the Pi uptake from soil is also greatly enhanced by the up-regulation of a high affinity Pi transport system in the presence of a lower environmental Pi concentration (Dunlop *et al.*, 1997). After the first membrane Pi transporters were identified in *Arabidopsis* and later designated as the Phosphate Transporter 1 (PHT1) family members, other PHT1 members have been identified in *Arabidopsis* and other plants (Muchhal *et al.*, 1996; Misson *et al.*, 2004; Nussaume *et al.*, 2011). Members of the PHT1 family are generally Pi / H^+ co-transporters that belong to the major facilitator superfamily (MFS) and are each predicted to share a similar topological structure, comprising 2 bundles of 6 trans-membrane α-helices and have both the C- and N-termini located on the cytosolic side of the membrane (Baker *et al.*, 2015). In *Arabidopsis*, only PHT1;1 and PHT1;4 are root specific Pi transporters and have been identified as homologues of yeast high-affinity Pi transporter Pho84, while the other seven PHT1 members are widely expressed in shoot and pollen (Shin *et al.*, 2004; Mudge *et al.*, 2002), indicating a plant-wide Pi uptake and remobilization function of the PHT1 family. The expression level of PHT1 genes is up-regulated by transcription factors PHR1 and WRKY75 while down-regulated by transcription factor MYB62 (Devaiah *et al.*, 2007; Devaiah *et al.*, 2009). The homeostasis of PHT1 is post-translationally modulated by

PHF1 (PHosphate transporter traffic Facilitator 1) through protein trafficking and membrane targeting (González *et al.*, 2005; Bayle *et al.*, 2011). The post-translational regulation of PHT1 also involves ubiquitination and subsequent endocytosis and vacuolar degradation by NLA (RING-type ubiquitin E3 ligase) and PHO2 (ubiquitin E2 conjugase) (Lin *et al.*, 2013), whereas NLA and PHO2 are themselves regulated by miRNAs that are induced by Pi starvation (Bari *et al.*, 2006; Hsieh *et al.*, 2009) (Figure 1.4; Table 1.1).

It is of note that due to the sessile nature of plants, formation of a mutualistic arbuscular mycorrhizal (AM) symbiosis can greatly increase the plant Pi uptake from soil (Figure 1.4). Around 80% of plant species form this type of association, but not brassicas which includes *Arabidopsis*. In this symbiosis system, the hyphal germ tube from germinated AM spores forms an appressorium upon contacting plant roots. The hyphae enter the root from the appressorium and develop around root cells before penetrating the plant root inner cortex cells. The arbuscules, a dichotomously branched structure, is formed from the hyphae inside the cortex cells (Parniske, 2004; Harrison, 2005). The peri-arbuscular membrane that generated by the plant cells around the arbuscules thus provides the symbionts nutrient exchange interface (Harrison, 2005; Javot *et al.*, 2007). Although the entire mechanism involved in nutrients interchange between plants and mycorrhizal fungus remains to be elucidated, specific PHT1 family members have been demonstrated to play key roles in this essential plant Pi uptake process. Among these PHT1 transporters, two subfamilies have been classified, namely the subfamily III Pi transporters whose expression is enhanced by symbiosis and the subfamily I Pi transporters whose expression is specifically seen in symbiosis (Javot *et al.*, 2007). Although physically connected to the plasma membrane of cortical cells, the periarbuscular membrane harbors a distinct population of proteins that are critical for AM symbiosis. Studies of *Medicago truncatula* MtPT4 has demonstrated establishment of this periarbuscular membrane specific protein pool is achieved by a transient expression and re-orientated secretion of Pi transporters in the colonized root cells (Pumplin *et al.*, 2012). Interestingly, studies in monocotyledon plant rice (*Oryza sativa*) showed that although the Pi demands of mycorrhizal rice could be solely satisfied by OsPT11, another phylogenetically distant PHT1 member OsPT13 is also required for the integrated formation of AM symbiosis, indicating non-redundant functions of AM-specific Pi transporters (Yang *et al.*, 2012).

1.4.2 Pi efflux and translocation

The maintenance of Pi homeostasis in plants requires specialized transporters not only for uptake of Pi from the soil but also for its cell-to-cell-movement within the plant, together with a rigorous regulation system for the functional integration of such transporters in the response to Pi starvation. The crucial role of the PHO1 family members has been recognized since the first isolation of the *pho1* mutant in *Arabidopsis.* This mutant shows deficient Pi accumulation in leaves as well as other Pi deficiency symptoms in the presence of adequate phosphate supply (Poirier *et al.*, 1991). Despite the greatly reduced leaf Pi content, Pi uptake rate into the roots and Pi transfer in the shoot xylem were not affected in *pho1* plants, indicating a specific function for PHO1 in Pi export from root cortical cells to the xylem for transport to the shoot (Figure 1.5) (Poirier *et al.*, 1991).

The gene encoding PHO1 was later cloned in a map-based strategy (Hamburger *et al.*, 2002). The *Arabidopsis* genome encodes another 10 homologues, each of which contains a conserved, hydrophilic N-terminal SPX domain (see section 1.5) and a hydrophobic C-terminal EXS (ERRD 1-XPR1-SYG1) domain with six to eight predicted membrane-spanning segments (Hamburger *et al.*, 2002). Although a broad expression pattern has been revealed for most of the *Arabidopsis* PHO1 homologues (Wang *et al.*, 2004), only PHO1, PHO1;H1 and PHO1;H10 expression is induced by Pi deficiency (Yuan and Liu, 2008). Only PHO1 and PHO1;H1 are found to be responsible for Pi export, and their activity is under the regulation of E2 conjugase PHO2 and the main transcription factor PHR1, respectively (Stefanovic *et al.*, 2007; Liu *et al.*, 2012) (Figure 1.4). Limited functional redundancy was also found between PHO1;H1 and PHO1, as the *pho1;h1* single mutant shows no detectable phenotypic changes compared to the wild type while the *pho1/pho1;h1* double mutant exhibits strong impairment both in growth and in Pi root-to-shoot transfer (Stefanovic *et al.*, 2007). This indicates the existence of complicated and potentially complementary regulatory pathways for Pi deficiency resistance.

The study of PHO1 family members in both *Arabidopsis* and rice has further verified their Pi export function as well as their responsibility for Pi transfer from roots to shoots (Stefanovic *et al.*, 2011; Secco *et al.*, 2010). Recent investigation of XPR1, the closest

14

mammalian homologue of PHO1, also demonstrated similar sub-cellular localization as well as Pi export mediating functions, though the N-terminal SPX domain has been shown to be dispensable for Pi efflux (Wege and Poirier, 2014; Giovannini *et al.*, 2013). However, in spite of the accumulating evidence showing a Pi export function for PHO1 and its homologues, it is still unclear how this function is modulated intracellularly in response to changes in Pi levels. Although translocation of PHO1 from the endomembrane system to the plasma membrane has been observed upon high Pi infiltration (Arpat *et al.*, 2012), increased levels of XPR1 at the metazoan cell surface did not lead to elevated Pi export, indicating that additional factors may be required, at least in mammalian cells (Giovannini *et al.*, 2013).

While PHO1 mediates efflux from the cytosol to the apoplast, a recent study in *Arabidopsis* revealed that the tonoplast localized Vacuolar Phosphate Transporter 1 (VPT1) is responsible for the Pi efflux from cytosol into the vacuole for storage (Liu *et al.*, 2015) (Figure 1.4). This finding has provided important information on plant vacuolar Pi sequestration and storage, whereas future work is needed to fully understand the mechanism of plant adaptation to variable Pi availability. One key step is to identify components that are involved in Pi release from the vacuole upon Pi starvation.

1.4.3   Integration of Pi sensing and signalling

The morphologic and genetic adaptations to Pi deprivation are essential in the maintenance of Pi homeostasis in plants and are regulated by complicated signalling pathways both locally and distantly between different plant tissues. Low Pi sensing triggers the regulation of multiple intracellular transcription factors (López-Arredondo *et al.*, 2014), among which the MYB family member Phosphate Starvation Response 1 (PHR1) plays a critical role of regulating downstream PSI genes by binding to a P1BS (PHR1-Binding Site) motif in their promoter regions (Rubio *et al.*, 2001). The nuclear localized PHR1 is responsible for the Pi starvation induced upregulation of Phosphate Transporter 1 (PHT1) family members, Pi exporter PHO1;H1, regulatory microRNA399 (miRNA399) as well as the expression level changes of SPX-exclusive family member AtSPX1-4 (see section 1.5.2.1) (Rubio *et al.*, 2001; Bari *et al.*, 2006; Secco *et al.*, 2012a)

(Figure 1.4; Table 1.1). Although PHR1 has previously been shown to be post-translationally controlled by SIZ1 through sumoylation (Miura *et al.*, 2005), recent studies in both *Arabidopsis* and rice (*Oryza sativa*) have revealed that the activity of AtPHR1/OsPHR2 is also modulated by SPX1/2 in a Pi dependent manner through a competitive inhibition of PHR1 binding to P1BS (Puga *et al.*, 2014; Wang *et al.*, 2014). Another study in rice also demonstrated under Pi sufficient condition, OsSPX4 can negatively regulate the OsPHR2 activity by retaining it in the cytoplasm as well as by interacting with the P1BS binding domain of nuclear localized OsPHR2 (Lv *et al.*, 2014). These findings, together with the most recently identified Pi/InsPs binding function of SPX domains (Wild *et al.*, 2016), have provided valuable insight into how plants sense Pi levels and integrate the signals into intracellular pathways, although further studies are needed to better understand the Pi signalling mechanisms on the molecular level. Long-distance signal transduction has also been demonstrated by split-root experiments in tomato (*Lycopersicon esculentum* L.), where the genetic Pi starvation responses in Pi deficient tissues can be repressed by Pi repletion in other distant tissues within the same plant (Liu *et al.*, 1998).

A mutant allele of the sucrose transporter gene SUC2 (*pho3*), also exhibits lower root acid phosphatase activity, low Pi content and other Pi-deficient symptoms along with a significant expression increase of the plastidic glucose 6-phosphate/phosphate translocator (Zakhleniuk *et al.*, 2001; Lloyd and Zakhleniuk, 2004). Mutations within the negative regulator of ethylene biosynthesis Ethylene Overproduction 1 (ETO1) have also been found to lead to the altered expression of Pi starvation induced (PSI) genes and enhanced acid phosphatase (APase) production independently of the Pi supply (Wang *et al.*, 2012b). These observations together demonstrate the involvement of other components such as sugar (Hammond and White, 2011) and phytohormones in the complicated signalling mechanisms of systemic Pi regulation in plants (Figure 1.4; Table 1.1).

Table 1.1 Profile of key elements involved in the regulation of Pi homeostasis in model plant *Arabidopsis* **[Modified from (Secco *et al.*, 2012a)]**

Protein Name	Function
AtLPR1	Multi-copper oxidase Regulates primary root cell elongation in response to low Pi
AtSIZ1	SUMO E3 ligase Facilitates the sumoylation of AtPHR1
AtPDR2	P5 type ATPase Maintain proper expression of scarecrow (SCR, regulator of root patterning) and maintain stem-cells in Pi-deprived roots
AtPHR1	MYB-CC family transcription factor Activates a subset of PSI genes Negatively regulated by AtSPX1/2 under Pi sufficient conditions
AtPHO1	Pi transfer from root to shoot; Pi loading into the xylem vessel Involved in Pi signalling
AtPHO1;H1	Pi transfer from root to shoot
AtPHO2	E2 conjugase Regulates Pi uptake, allocation and remobilization
AtPHO3	Mutation has less acid phosphatase (APase) activity in roots but higher APase activity in shoots; mutants also accumulate less P content but higher anthocyanin and carbohydrates content
AtETO1	Negative regulator of ethylene biosynthesis Mutation causes altered expression of PSI genes and enhanced production of APase
AtVPT1	Pi efflux from cytosol into the vacuole for storage
AtNLA	RING-type ubiquitin E3 ligase Regulates Pi homeostasis under N limitation condition
AtPHF1	SEC12-related protein PHT1 trafficking facilitator
AtPHT1	High affinity Pi transporter
AtSPX1 /2	Positive regulator of plant adaptation to Pi starvation Interacting with PHR1 in a Pi dependent manner
AtSPX3	Negative regulator of some PSI genes
WRKY75	WRKY transcription factor Regulates nutrient starvation response and root development
MYB62	R2R3-type MYB transcription factor Regulates phosphate starvation responses via changes in GA metabolism and signalling

1.5 The conserved SPX domain and SPX domain-containing proteins in Pi regulation

The understanding of plant responses to Pi starvation has been increased by a number of newly discovered key components in the Pi sensing and signalling pathways. Many of these key players, including the previously discussed PHO1 family members, have been found to contain an SPX domain (Hamburger *et al.*, 2002; Duan *et al.*, 2008), a conserved protein domain that appears to be heavily involved in the maintenance of Pi homeostasis in various organisms.

The SPX domain, originally found in the yeast *Saccharomyces cerevisiae*, is named after the Suppressor of Yeast gpa1 (Syg1), the yeast Phosphatase 81 (Pho81) and the human Xenotropic and Polytropic Retrovirus receptor 1 (Xpr1) (Secco *et al.*, 2012a). While the C-terminal regions of SPX-domain proteins can differ greatly in their nature and organization, the SPX domain always constitutes a well conserved N-terminus in these proteins. The SPX domain varies in length from 135-380 amino acids, depending on the arrangement of the low similarity insertion sequences that separate the three well conserved sub-domains, each of which consists of 30-40 amino acids (Secco *et al.*, 2012a) (Figure 1.6).
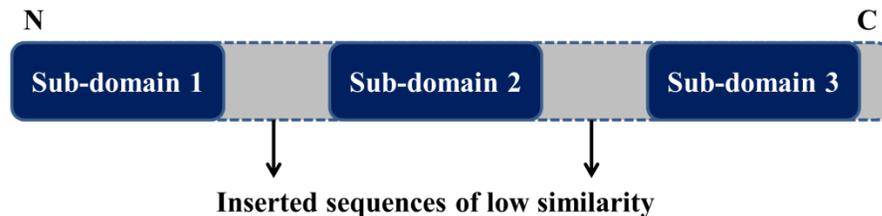
Figure 1.6 Schematic representation of the SPX domain. [Modified from (Secco *et al.*, 2012a)]

The blue boxes indicate three well-conserved sub-domains (sub-domains 1-3), separated from each other by low similarity regions of various lengths which are symbolized by the dotted line.

1.5.1 SPX domain-containing proteins in yeast

In yeast, a frequently investigated eukaryotic 'model', eight out of the ten proteins that possess the SPX domain are found to be involved in the maintenance of Pi homeostasis. The first to be characterized, Pho81, a cyclin-dependent kinase (CDK) inhibitor, has been well demonstrated as a negative regulator in the phosphate-responsive signalling pathway (section 1.3 and Figure 1.3). Three low-affinity Pi transporters, Pho87, Pho90 and Pho91 (section 1.3 and Figure 1.3), are also members of the yeast SPX domain-bearing protein group. Although distinct vacuolar targeting pathways have been detected for Pho87 and Pho90, either dependent or independent on the regulatory protein Spl2, respectively, the N-terminal SPX domain is strictly required for the vacuolar targeting process (Ghillebert *et al.*, 2011), suggesting a potential protein-protein interaction function of the SPX domain.

The second SPX domain-containing group of proteins in yeast is from the Vacuolar Transporter Chaperone (Vtc) family. Although four members from this family are known to be up-regulated by Pi starvation, only three of them, namely Vtc2, Vtc3 and Vtc4, possess the SPX domain at their N-termini (Ogawa *et al.*, 2000). The vacuolar inorganic polyphosphate (polyP) accumulation has been demonstrated to rely strictly on the formation of the integrated Vtc complex, which consists both Vtc1 and Vtc4 and either Vtc2 or Vtc3 (Ogawa *et al.*, 2000). This suggests protein-protein interactions among these Vtc family members, possibly via the SPX domain, during the assembly of Vtc complexes. Although Vtc4 was shown to be a polyP synthesizing enzyme (Hothorn *et al.*, 2009), the N-terminal SPX domain of Vtc4 was absent in this functional study, showing that the SPX domain is not necessarily involved in the catalytic activity of the enzyme. Membrane trafficking and signal transduction functions have also been proposed for the SPX domain in the vacuolar targeting of Vtc2 upon Pi stress (Hothorn *et al.*, 2009).

A further SPX domain-containing protein in yeast is GDE1 (glycerophosphocholine phosphodiesterase1), the only glycerophosphodiester phosphodiesterase identified in yeast to hydrolyze glycerophosphocholine into glycerophosphate and choline under low Pi conditions (Fisher *et al.*, 2005). In contrast to its *in vivo* function, the *in vitro* phosphodiesterase activity of GDE1 has not been detected so far, raising the hypothesis that its biological activities require physical interaction with other partners (Fisher *et al.*, 2005), which can possibly be achieved by protein interactions via the SPX domain.

However, other glycerophosphodiester phosphodiesterases involved in plant Pi homeostasis maintenance, have been shown to lack an SPX domain (Cheng *et al.*, 2011). This again suggests the SPX domains may not be necessarily required for the *in vivo* catalytic activities of these enzymes.

1.5.2 SPX domain-containing proteins in plants

Many of the SPX domain-containing proteins in yeast belong to the PHO regulons, which have not been found in plants. Nevertheless, the SPX domain is still found in many plant proteins involved in Pi signalling (Table 1.2), and these proteins can be sub-divided into four groups based on the presence or absence, and nature, of their appended C-terminal domains: the SPX exclusive, SPX-EXS, SPX-MFS and SPX-RING families (Secco *et al.*, 2012a).

1.5.2.1 SPX exclusive family

In plants, members of the SPX exclusive family of proteins contain the SPX domain as the only identified protein domain, while their C-terminal regions are variable in sequence and contain segments without defined secondary structure. In *Arabidopsis*, four small proteins in this family with an average length of 280 amino acids have so far been identified and designated AtSPX1-AtSPX4 (Secco *et al.*, 2012a). In response to Pi starvation, expression levels of AtSPX1, AtSPX2 and AtSPX3 are significantly enhanced while the transcript level of AtSPX4 is reduced to half of normal levels (Duan *et al.*, 2008). Transgenic plants deficient in AtSPX1, AtSPX2 and AtSPX4 showed no detectable phenotype in conditions of Pi sufficiency, suggesting possible *in vivo* functional compensation among these SPX exclusive family members. Over-expression of AtSPX1 increases the expression of several Pi-responsive genes, including ACP5 (Acid Phosphatase 5), RNS1 (Ribonuclease 1) and PAP2 (Purple Acid Phosphatase 2). This, together with its nuclear localization (Duan *et al.*, 2008), suggests a transcriptional regulator function of AtSPX1 in response to Pi starvation. Repression of AtSPX3 by RNAi resulted in a hypersensitive response to Pi starvation, with enhanced expression of Pi-signalling genes in roots and Pi-recycling genes in shoots, respectively (Duan *et al.*, 2008), leading to increasing shoot Pi concentrations and decreasing root Pi concentrations until the plant stopped growing. This indicates the involvement of AtSPX3 in a negative feedback loop in Pi starvation responses. Similar Pi signalling pathway components were subsequently identified in rice as OsSPX1-OsSPX6 (Wang *et al.*, 2009b). Functional similarity with AtSPX3 has been demonstrated for OsSPX1 as a negative regulator (Wang *et al.*, 2009a).

Although the expression of SPX exclusive family members has previously been shown to be under the control of the central transcription factor PHR1 and its homologues (Duan *et al.*, 2008; Wang *et al.*, 2009a), recent studies have also demonstrated a Pi-dependent interaction between SPX exclusive proteins and their molecular regulators (Puga *et al.*, 2014; Wang *et al.*, 2014) (see section 1.4.3). This interaction thus generates a negative feedback loop on the expression control of these SPX exclusive family members and suggests a tight transcriptional control of these proteins under Pi sufficient conditions.

1.5.2.2   SPX-EXS family

The PHO1 proteins are currently the only identified plant SPX-EXS family, containing a hydrophobic EXS domain at their C-termini (Wang *et al.*, 2004). The name EXS comes from the yeast protein ERD1 (Endoplasmic Reticulum Retention Defective 1), the human protein XPR1 and the yeast protein SYG1. No established function has so far been defined for the EXS domain despite its wide occurrence in eukaryotes. The inserted low similarity regions in the SPX domain found in this family are much longer than the ones in other SPX domain-containing protein families (Secco *et al.*, 2012a). In *Arabidopsis*, besides the critical roles of AtPHO1 and AtPHO1;H1 in Pi signalling pathways discussed above (see section 1.4.2), the homologue AtPHO1;H4 has been shown to be involved in control of hypocotyl elongation under blue light (Zhou and Ni, 2010), while AtPHO1;H10 has been shown to be involved in the responses to a number of biotic and abiotic stresses other than Pi starvation (Ribot *et al.*, 2008). However, no other protein member involved in Pi sensing or signalling has been identified within this SPX-EXS family.

In rice, only three members, OsPHO1;1-OsPHO1;3, have been identified in the PHO1 family. Phylogenetic study shows that all three of these proteins belong to the same clade as the two *Arabidopsis* Pi transporters AtPHO1 and AtPHO1;H1 (Secco *et al.*, 2012a). To date, only OsPHO1;2 has been well demonstrated to be responsible for Pi transfer from roots to shoots and this transferring function is possibly regulated by its cis-Natural Antisense Transcripts (NATs), which are highly induced upon Pi starvation (Secco *et al.*, 2010).

1.5.2.3 SPX-MFS family

The Major Facilitator Superfamily (MFS) is well known as one of the largest and almost ubiquitous families of transporters for small molecules in both prokaryotes and eukaryotes (Pao *et al.*, 1998). Presence of a C-terminal MFS domain in members of the SPX-MFS protein family likely confers upon them the putative molecule transporting and signalling functions. So far, three SPX-MFS family members have been identified in rice, and shown to exhibit preferential expression in the shoot (Lin *et al.*, 2010). Quantitative RT-PCR analysis revealed that Pi starvation suppresses the expression of OsSPX-MFS1 and OsSPX-MFS3 but increases the transcription level of OsSPX-MFS2 (Lin *et al.*, 2010). In addition, a key role has later on been established for OsSPX-MFS1 in maintaining Pi homeostasis in the leaves, potentially involving its function as a Pi transporter (Wang *et al.*, 2012a). Despite the opposite transcriptional responses to Pi starvation, both OsSPX-MFS1 and OsSPX-MFS2 have been identified to be negatively regulated by osa-miR827 (Lin *et al.*, 2010), suggesting distinct and complex regulation mechanisms of the SPX-MFS proteins.

In *Arabidopsis*, sequence comparison of miR827 and target genes identified a SPX-MFS protein encoding gene At1g63010 as a potential target of ath-miR827 (Lin *et al.*, 2010). The homologues of *Arabidopsis* SPX-MFS proteins in barley were identified as vacuolar proteins (Endler *et al.*, 2006), and very recently, the protein encoded by At1g63010 was demonstrated as a vacuolar Pi transporter and designated as VPT1 (Liu *et al.*, 2015). VPT1 was found to be responsible for the vacuolar Pi accumulation primarily in younger tissues while its expression in older tissues is enhanced by increasing environmental Pi levels (Liu *et al.*, 2015). *Vpt1* mutants were also found to be hypersensitive to Pi starvation conditions due to the lower vacuolar Pi level (Liu *et al.*, 2015).

1.5.2.4 SPX-RING

Members of a fourth family of SPX domain-containing proteins are known to bear a Really Interesting New Gene (RING) finger domain at their C-termini. The RING domain, a characteristic feature of the largest family of E3 ubiquitin ligases, binds two zinc ions and plays an essential role in ubiquitin transfer by protein interaction with ubiquitin E2

conjugase (Budhidarmo *et al.*, 2012). Although RING domains have been discovered in many proteins, AtNLA (Nitrogen Limitation Adaptation) is the only gene encoding an SPX-RING protein that has been characterized in *Arabidopsis* in connection with nutrient stress (Peng *et al.*, 2007). AtNLA was first identified to be involved in *Arabidopsis* adaptation to N limitation: the *Atnla* mutant cannot adapt to low N supply, exhibiting a decreased accumulation of anthocyanins and consequent early senescence (Peng *et al.*, 2007). Interestingly, this fatal phenotype could be rescued by reducing the Pi supply, thus triggering another anthocyanin synthesis pathway independent of AtNLA (Peng *et al.*, 2007), or by suppressing AtPHF1 and AtPHT1;1, the possible targets of AtNLA and AtPHO2 (Kant *et al.*, 2011). In addition, it has been shown that AtNLA expression is regulated by miR827 which is induced upon low Pi concentration, and that the early senescence in the *Atnla* mutant is caused by the toxic high Pi content found in shoots grown under a low N background (Kant *et al.*, 2011). This phenomenon resembles the *pho2* mutant-induced high shoot Pi toxicity (Delhaize and Randall, 1995). Given the fact that the E2 conjugase AtPHO2 and E3 ligase AtNLA are both ubiquitination pathway components and negative regulators of shoot Pi content, these two proteins can possibly participate together in controlling the nitrate-dependent Pi homeostasis. Furthermore, interaction between the N-terminal SPX and C-terminal RING domains could possibly be involved in the regulation of SPX-RING family proteins, as protein interaction has been previously found in the RING domain of AtNLA (Peng *et al.*, 2007).

Table 1.2 Profile of SPX domain-containing proteins in *Arabidopsis* (At) and rice (*Oryza sativa*, Os). [Modified from (Secco *et al.*, 2012a)]

	Protein	Function/Regulation profile	Subcellular localization
SPX exclusive	AtSPX1	Positive regulator of plant adaptation to Pi starvation Interacting with PHR1 in a Pi dependent manner	Nucleus
	AtSPX2	Interacting with PHR1 in a Pi dependent manner	Nucleus
	AtSPX3	Negative regulator of some PSI genes	Cytoplasm speckles
	AtSPX4	-	Cytoplasmic membrane
	OsSPX1	Positive regulator of plant adaptation to Pi starvation Interacting with PHR2 in a Pi dependent manner	Nucleus
	OsSPX2	Interacting with PHR2 in a Pi dependent manner	Nucleus
	OsSPX3	Negative regulator of some PSI genes	Cytoplasm speckles
	OsSPX4	Interacting with PHR2 mainly in cytoplasm and preventing its translocation into the nucleus Controlled by Pi dependent 26S Proteasome Pathway	Nucleus/Cytoplasm
	OsSPX5	-	-
	OsSPX6	-	-
SPX-EXS	AtPHO1	Pi transfer from root to shoot; Pi loading into the xylem vessel Possible transcriptional signal transporting from root to shoot. Controlled by PHO2 mediated endomembrane degradation	Largely at Golgi/trans-Golgi network and uncharacterized vesicles; A minor fraction at plasma membrane
	AtPHO1;H1	Pi transfer from root to shoot Regulated by PHR1 and influenced by phosphite	-
	AtPHO1;H2	-	-
	AtPHO1;H3	-	-
	AtPHO1;H4 (AtSHB1)	Control hypocotyl elongation under blue light Form a large protein complex through SPX and EXS domain Regulate endosperm development relevant genes	Nucleus
	AtPHO1;H5	-	-
	AtPHO1;H6	-	-
	AtPHO1;H7	-	-
	AtPHO1;H8	-	-

	AtPHO1;H9	-	-
	AtPHO1;H10	Involved in abiotic/biotic stresses response (including wounding, dehydration, cold, salt and pathogen attack)	-
	OsPHO1;1	-	-
	OsPHO1;2	Pi transfer from root to shoot Gene expression regulated by its cis-natural antisense transcripts	-
	OsPHO1;3	-	-
SPX-MFS	AtSPX-MFS1	-	-
	AtSPX-MFS2	-	-
	AtSPX-MFS3	-	Tonoplast
	OsSPX-MFS1	Pi transport and relocation in leaves Gene expression controlled by miR827	-
	OsSPX-MFS2	Gene expression controlled by miR827	-
	OsSPX-MFS3	-	-
	OsSPX-MFS4	-	-
SPX-RING	AtNLA (AtBAH1)	Involved in the nitrogen starvation response Regulating Pi homeostasis under nitrogen limited condition	Nuclear speckles
	AtNLA2	-	-
	OsNLA1	-	-
	OsNLA2	-	-

1.5.3 Structural characterization of SPX domain

Since the first identification of SPX proteins in yeast, the sequence conservation of the SPX domain among all major eukaryote groups and its involvement in plant responses to various biotic and abiotic stresses have been well recognized (Secco *et al.*, 2012a; Peng *et al.*, 2007; Lin *et al.*, 2010). However, though new interacting partners of SPX protein have been identified, the exact Pi-binding function of this highly conserved SPX domain was not confirmed until the very determination of its molecular structure (Wild *et al.*, 2016) (Figure 1.7).
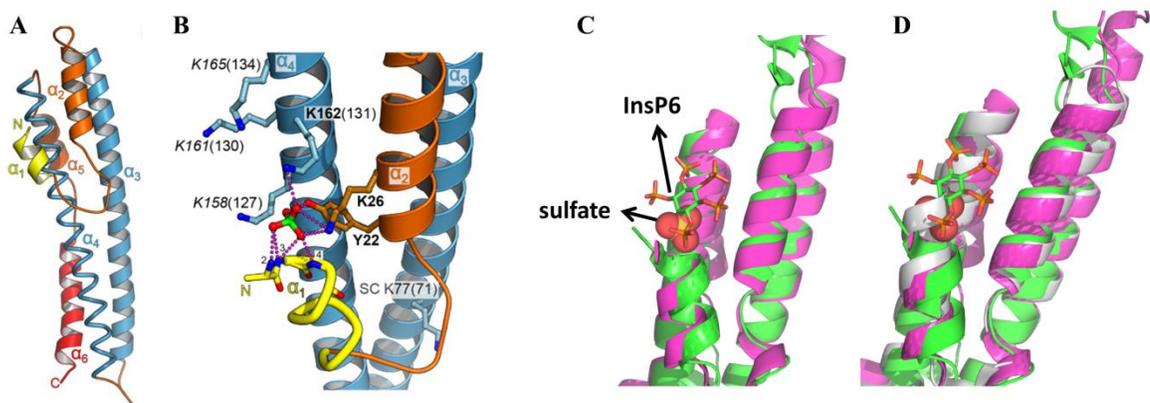
Figure 1.7 Structure of SPX domains and the basic surface cluster (Wild *et al.*, 2016).

Crystal structures of SPX domain from ScVtc4 (*Saccharomyces cerevisiae*), CtGde1 (*Chaetomium thermophilum*) and HsXPR1 (*Homo sapiens*) were solved at 2.1 - 3.0 Å, 1.95 Å and 2.43 Å resolution, respectively by Wild *et al.*. A. ScVtc4 SPX domain structure without ligand binding (Wild *et al.*, 2016); B. HsXPR1 SPX domain structure with sulfate ion binding to the basic surface cluster (Bold characters: Phosphate Binding Cluster; Italics characters: Lysine Surface Cluster; SC: a distant Lys residue; Bracketed numbers: ScVtc2 corresponding residues) (Wild *et al.*, 2016); C. Overlaid SPX domain N-terminal structures of CtVtc4 with InsP6 bound (Green, Protein data bank code 5IJP) and HsXPR1 with sulfate ion bound (Magenta, Protein data bank code 5IJH); D. Overlaid SPX domain N-terminal structures of CtVtc4 with InsP6 bound (Green, Protein data bank code 5IJP), HsXPR1 with sulfate ion bound (Magenta, Protein data bank code 5IJH) and ScVtc4 without ligand binding (Grey, Protein data bank code 5IIG).

Crystallographic data of the SPX domain from various eukaryotic organisms has shown a six-α helical structure. The C-terminal α helix is demonstrated to have different orientations in SPX domains from different species while the N-terminus is disordered without ligand binding (Wild *et al.*, 2016) (Figure 1.7 A, D). Not only was a Phosphate Binding Cluster (PBC) identified in SPX domains for its Pi-binding capacity, it was also found to form a basic ligand binding surface with a Lysine Surface Cluster (KSC) for interacting with other Pi-containing ligands, such as InsPs with much higher affinities (μM range) (Wild *et al.*, 2016) (Figure 1.7 B, C). Mutations in the conserved residues within this basic binding surface strongly reduced the SPX domain ligand binding capacity *in vitro* and also interfere with the *in vivo* Pi signalling pathways (Wild *et al.*, 2016). Overlaying the SPX domain N-terminal structures with and without ligand binding showed the N-terminus of α1 helix occupies the position of Pi or Pi-containing ligands in the apo structure (Figure 1.7 D). During ligand interaction, position of α1 helix shifts downwards to accommodate the ligand (Figure 1.7 C, D). This ligand binding induced conformational change could also help to stabilized the otherwise disordered N-termini of SPX domains (Wild *et al.*, 2016).

Earlier *in vitro* studies have shown Pi at mM range could promote the interaction between *Arabidopsis* and rice SPX exclusive members and PHR1 transcription factors (Puga *et al.*, 2014; Wang *et al.*, 2014). However, the same interaction between these SPX and PHR1 proteins was found to be triggered by a much lower concentration of InsPs (Wild *et al.*, 2016). This, together with the previously known interaction between yeast Pho81 and IP7 during Pi starvation response (Lee *et al.*, 2008), indicates InsPs instead of Pi are the Pi signalling molecules *in vivo*. Therefore, the Pi-containing ligand sensing and binding capacity of SPX domain is likely to put the SPX domain-containing proteins in an intermediate position of passing on the intracellular Pi signals and interacting with other proteins in multiple Pi signalling cascades in eukaryotic cells.

1.6 Project Aims

The primary focus of this project is to characterize the SPX exclusive family members in plant Pi sensing and regulation. Although proteins from this family have been intensely studied in model plant *Arabidopsis*, no study has been conducted using the agriculturally important plant potato (*Solanum tuberosum*). Therefore this project aims at identifying the SPX domain-containing proteins in potato as well as to characterize the roles of StSPX exclusive family members during potato Pi starvation responses. Another area of focus from this study was to solve the structure of the conserved SPX domain. However, during the process of the course, the structures of SPX domain were made available, confirming its Pi-containing ligand binding capacity whereas investigations in model plant *Arabidopsis* have also demonstrated the interaction between SPX exclusive family members and transcription factor PHR1. Nevertheless, in spite of these latest advances, the molecular mechanism of AtSPX1-AtPHR1 interaction is still unclear. This project, therefore, also aims to better understand the working mechanism of AtSPX1-AtPHR1 interaction and its effect on the regulation of PSI gene expression and intracellular Pi level.

The results presented here present a novel working model for the molecular interacting mechanism between AtSPX1 and AtPHR1 and provide the first evidence of the important roles played by SPX proteins in the Pi starvation response in crop plant potato. These findings on how plant sense and integrate the intracellular Pi signals through these SPX exclusive family members also provide valuable insight into the design and cultivation of crop plants with a better Pi use efficiency.

Chapter 2 Materials and Methods

2.1 Materials

2.1.1 Reagents

Analar grade reagents were purchased from Sigma-Aldrich, Melford, Promega, Thermo Fisher, GE Healthcare and Fisher Scientific.

2.1.2 Water

ASTM Type I water (purified and deionized water, resistivity 18.2 Megaohm/cm) produced by ASTM Type I Ultra Pure Water Systems (Sartorius) was used in all solution preparations unless otherwise stated. Sterilization was carried out by autoclaving at 121°C for over 20 min or filtering through 0.22 μm filters where required.

2.1.3 Standard solutions

All solutions were made up with water described above unless otherwise stated. The composition of common solutions is listed in this section and other buffer composition is included in the experimental procedure section.

2.1.4 Enzymes for molecular biology

Restriction enzymes were purchased from New England Biolabs and used with corresponding buffers recommended and supplied by the manufacturer.

KOD hot start DNA polymerase (Novagen), Precision nanoScript2 Reverse Transcriptase (Primerdesign) and T4 DNA ligase (New England Biolabs) were purchased and used according to the manufacturers' instructions.

2.1.5 DNA size markers

1 kb, 100 bp DNA Ladders and 6x agarose gel loading dye were purchased from New England Biolabs and used according to the manufacturer's instructions.

2.1.6 Protein size markers

PageRuler Prestained Protein Ladder – 10 to 180 kDa (Thermo Fisher) and Precision Plus Unstained Protein Standards (Bio-rad) were purchased and used according to the manufacturers' instructions.

2.1.7 Manufacturers' complete systems

Wizard Plus SV Minipreps DNA Purification System (Promega), RNeasy Plant Mini Kit (Qiagen) and Precision nanoScript 2 Reverse Transcription kit (Primerdesign) were used according to the manufacturers' instructions.

2.1.8 Protease inhibitors

Complete EDTA-free Protease Inhibitor Cocktail (Roche) and Protease Inhibitor Cocktail For Plant Cell Extracts (Sigma-Aldrich) were used according to the manufacturers' instructions.

2.1.9 Protein analysis solutions

2.1.9.1 Cell lysis

10x Phosphate-buffered saline (PBS): Na_2HPO_4 (100 mM), KH_2PO_4 (18 mM), NaCl (1370 mM), KCl (40 mM).

Enzyme lysis solution: HEPES (50 mM), $MgCl_2$ (5 mM), Triton X-l00 (1% v/v), sucrose (25% w/v), OmniCleave endonuclease (10 U/mL), lysozyme (0.1 mg/mL), pH 8.0.

2.1.9.2 SDS-PAGE

4 x separating buffer: Tris-HCl (1.5 M), SDS (0.4% w/v), pH adjusted to 8.8 with HCl.

4 x stacking buffer: Tris-HCl (0.5 M), SDS (0.4% w/v), pH adjusted to 6.8 with HCl.

Table 2.1 Composition of 12% SDS-PAGE gels (volumes given are sufficient for two 0.75 mm gels or one 1.5 mm gel)

Component	12% Separating gel	4% Stacking gel
Acrylamide/ Bis acrylamide ratio= 37.5:1 (30% w/v solution)	6.0 mL	0.65 mL
4 x separating buffer	3.75 mL	-
4 x stacking buffer	-	1.25 mL
MilliQ H$_2$O	5.25 mL	3.05 mL
10% Ammonium persulphate (APS)	50 μL	25 μL
Tetramethylethylenediamine (TEMED)	10 μL	5 μL

10x SDS Running buffer: glycine (190 mM), Tris-HCl (25 mM), SDS (0.1% w/v). Diluted to 1x before use.

6x SDS Sample buffer: glycerol (50% w/v), Tris-HCl (0.5 M, pH 6.8), SDS (10% w/v), EDTA (10 mM), Pyronin Y (0.1% w/v), DTT (60 mM). Diluted to 1x before use.

2.1.9.3 Western blotting

Transfer buffer: Tris-HCl (25 mM), glycine (192 mM), methanol (20% v/v), kept at 4 ºC.

Tris-buffered saline Tween (TBST): Tris-HCl (50 mM), NaCl (150 mM), Tween 20 (0.1% v/v), pH 7.5.

Blocking buffer: Bovine serum albumin (BSA) (3% w/v) in TBST.

2.1.9.4 Coomassie Staining

Fixing solution: Acetic acid (10% v/v), isopropanol (25% v/v).

Staining solution I: Acetic acid (10% v/v), isopropanol (25% v/v), Coomassie Brilliant Blue R250 (0.025% w/v).

Staining solution II: Acetic acid (10% v/v), isopropanol (10% v/v), Coomassie Brilliant Blue R250 (0.0025% w/v).

Destaining solution: Acetic acid (10% v/v).

2.1.10 DNA analysis solutions

50x Tris-acetate-EDTA (TAE) buffer (per L): Tris-HCl (48.4 g), EDTA (2.92 g), acetic acid (11.4 mL). Diluted to 1x before use.

10x TE buffer: Tris-HCl (0.1 M), EDTA (10 mM), pH 8.0. Diluted to 1x before use.

2.1.11 Culture media

2.1.11.1 Standard media for IPTG induction

Lysogeny broth (LB) medium: Tryptone (1% w/v), yeast extract (0.5% w/v), NaCl (1% w/v), pH adjusted to 7.4 before autoclaving at 121°C for over 20 min.

Super broth (SB) medium: Tryptone (3.2% w/v), yeast extract (2% w/v), NaCl (0.5% w/v), pH adjusted to 7.4 before autoclaving at 121°C for over 20 min.

2.1.11.2 Media for auto-induction (Studier, 2005)

LB auto-induction medium: Tryptone (1% w/v), yeast extract (0.5% w/v), pH adjusted to 7.4 before autoclaving at 121°C for over 20 min.

SB auto-induction medium: Tryptone (3.2% w/v), yeast extract (2% w/v), pH adjusted to 7.4 before autoclaving at 121°C for over 20 min.

M9 auto-induction medium: Na_2HPO_4 (0.6% w/v), KH_2PO_4 (0.3% w/v), NH_4Cl (0.1% w/v), casamino acids (0.1% w/v), $CaCl_2$ (0.0003% w/v), sterilized by autoclaving at 121°C for over 20 min.

Complete media for auto-induction (M9 auto, LB auto and SB auto) (per L): To 929 mL auto induction medium add, in the following order, 1 mL 1 M $MgSO_4$, 20 mL 50 x 5052 (25 % (w/v) glycerol, 2.5 % (w/v) glucose, 10 % (w/v) α-lactose monohydrate) and 50 mL 20 x NSPC (0.5 M Na_2HPO_4, 0.5 M KH_2PO_4, 0.1 M Na_2SO_4, 1 M NH_4Cl).

2.1.11.3 Agrobacterium media

YEP medium: Tryptone (1% w/v), yeast extract (1% w/v), NaCl (0.5% w/v), pH adjusted to 7.0 before autoclaving at 121°C for over 20 min.

2.1.11.4 Antibiotics

When required, antibiotics were sterilized by filtering through 0.22 μm filters and added to media at the following concentrations: Ampicillin/ Carbenicillin (100 μg/mL), Kanamycin (50 μg/mL), Chloramphenicol (30 μg/mL), Rifampicin (50 μg/mL), Streptomycin (30 μg/mL).

2.1.12 Plant hydroponic nutrient solution (Hammond *et al.*, 2011)

High Pi (300 μM) nutrient solution: $Ca(NO_3)_2$ (2 mM), NH_4NO_3 (2 mM), $MgSO_4$ (0.75 mM), KOH (0.5 mM), FeNaEDTA (0.1 mM), $CaCl_2$ (25 μM), H_3BO_3 (30 μM), $MnSO_4$ (10 μM), $CuSO_4$ (3 μM), $ZnSO_4$ (1 μM), Na_2MoO_4 (0.5 μM), KH_2PO_4 (0.3 mM), pH 6.0.

Low Pi (10 μM) nutrient solution: $Ca(NO_3)_2$ (2 mM), NH_4NO_3 (2 mM), $MgSO_4$ (0.75 mM), KOH (0.5 mM), FeNaEDTA (0.1 mM), $CaCl_2$ (25 μM), H_3BO_3 (30 μM), $MnSO_4$ (10 μM), $CuSO_4$ (3 μM), $ZnSO_4$ (1 μM), Na_2MoO_4 (0.5 μM), KH_2PO_4 (0.01 mM), K_2SO_4 (0.145 mM), pH 6.0.

2.1.13 Antibodies

Primary

Polyclonal anti-Glutathione-S-Transferase (GST) antibody produced in rabbit (Sigma-Aldrich, G7781) was used at a 1:2000 dilution.

Anti-GFP rabbit polyclonal serum (Thermo Fisher. A6455) was used at a 1:2000 dilution.

Horseradish peroxidase-conjugated monoclonal anti-His$_6$ antibody (R&D Systems, MAB050H) was used at a 1: 5000 dilution.

Secondary

Horseradish peroxidase-conjugated secondary goat anti-rabbit antibody (Jackson Immuno Research Lab, 111-035-003) was used at a 1:20000 dilution.

2.1.14 Bacterial strains

2.1.14.1 Escherichia coli (E. coli) strains

OmniMAX™ 2 T1R (Invitrogen): F´ [*pro*AB+ *lac*Iq *lac*ZΔM15 Tn10(TetR) Δ(*ccd*AB)] *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *end*A1 *rec*A1 *sup*E44 *thi*-1 *gyr*A96 *rel*A1 *ton*A *pan*D

Lemo21 (DE3) (NEB): *fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS/ pLemo*(CamR) *λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5* pLemo = pACYC184-*PrhaBAD-lysY*

BL21 Star™ (DE3) (Invitrogen): F- *ompT hsdS*B(rB- mB-) *gal dcm rne131*(DE3)

BL21 Gold™ (DE3) (Stratagene): F- *dcm*+ Hte *omp*T *hsd*S(rB- mB-) *gal* λ (DE3) *end*A *Tet*r

C41 (DE3) (Lucigen Corporation): F– *omp*T *gal dcm hsd*SB(rB- mB-) (DE3)

C43 (DE3) (Lucigen Corporation): F– *omp*T *gal dcm hsd*SB(rB- mB-) (DE3)

2.1.14.2 Agrobacterium Strain

Agrobacterium tumefaciens LBA4404 strain was provided by Dr. Antony Ceasar Stanislaus and used for the AtSPX1 protein transient expression studies in tobacco plants.

2.1.15 Plant material

Seed potatoes (*Solanum tuberosum*, variety: Estima) were provided by Wellhouse Farm, Micklefield, UK. All *Arabidopsis thaliana* lines are of the *Columbia* ecotype. Tobacco seeds (*Nicotiana benthamiana*) were provided by Dr. Antony Ceasar Stanislaus.

2.1.16 Plasmids

2.1.16.1 Bacterial

pET28-b: Kan^R (Novagen, https://www.addgene.org/vector-database/2566/).

pL53: Amp^R (provided by Prof. Stephen Baldwin, Appendix 3).

pMAL-c2x: Amp^R (New England Biolabs, https://www.addgene.org/vector-database/3509/).

pGEX-2T: Amp^R (GE Healthcare, https://www.addgene.org/vector-database/2868/).

pGEX-4T-2: Amp^R (GE Healthcare, https://www.addgene.org/vector-database/2877/).

2.1.16.2 Plant binary vectors

pEAQ-*HT*: Kan^R bacteria, Kan^R plants (Provided by Prof George Lomonossoff, John Innes Centre, Norwich) (Sainsbury *et al.*, 2009).

2.2 Methods

2.2.1 Plant growth, transformation and analysis

2.2.1.1 *Potato hydroponic system*

Seed potatoes were germinated at 20°C and then grown under conditions of 20°C constant and 16 h light per day, using a hydroponic system based on that of Hammond and colleagues (Hammond *et al.*, 2011). Briefly, germinated seed potatoes were moved into perlite and supplied with deionized water before the 5-10 cm plantlets were transferred into the hydroponic system that was maintained in a greenhouse at 20°C and 16 h daylight. Once transferred to the hydroponic system, the plants were supplied with high Pi (300 μM) nutrient solution (section 2.1.12). Nutrient solutions were adjusted to pH 6.0, using H_2SO_4 daily and solutions were replaced completely twice a week. After the plants had adapted to the hydroponic system (7 days), Pi deprivation condition was introduced to half of the experimental plants by replacing the high Pi (300 μM) nutrient solution with low Pi (10 μM) nutrient solution (section 2.1.12). Plant tissues were sampled at the same time point (4pm) on the 1st, 7th, 14th, 21st and 28th days of treatment application.

For gene expression analysis, leaf tissue (the second leaf below the crown of the plant) and root tissue (2-3 cm from root tips) from high (300 μM) /low (10 μM) Pi treatments were sampled and immediately submerged in at least 10 volumes (approximately 10 μL reagent per 1 mg tissue) of RNA*later* RNA Stabilization Reagent (Qiagen, 76106). Tissues were preserved in RNA*later* RNA Stabilization Reagent and kept at -20°C before RNA extraction.

For plant tissue P content measurement, shoot and root tissues from hydroponically grown potato plants were harvested and oven-dried at 65-70ºC for 72 h before sample dry weights were recorded. Dried samples were subsequently milled for further P content measurement (section 2.2.2).

2.2.1.2 *Arabidopsis protoplast preparation and transformation*

Arabidopsis seeds were germinated in compost and grown in greenhouse under long day conditions (16 h light) at 21ºC. Leaves from 3 to 5-week-old plants were collected and protoplasts were isolated and transfected with 0 – 50 μg plasmid DNA, as described in Wu *et al.*, 2009. Excitation wavelength of 458 nm for GFP was provided by an X-Cite 120Q Wide-Field Fluorescence Microscope Excitation Light Source (Lumen Dynamics) and the expression of GFP tagged protein was examined with 480-520 nm emission band filter using Olympus BX61 Motorized System Microscope.

2.2.1.3 *Agrobacterium infiltration mediated transient protein expression in tobacco leaves*

<u>*Agrobacterium* transformation:</u> 1 μg of pEAQ-*HT*-AtSPX1-His$_6$ and pEAQ-*HT*-His$_6$-AtSPX1 plasmid DNA was added to 0.1 mL completely thawed *Agrobacterium* LBA4404 competent cells and mixed gently. Cells were then snap frozen in liquid nitrogen before being thawed at 37°C for 5 min. 150 μL YEP medium was added to the cells and incubated at 28°C for 2~4 hours with gentle shaking. After incubation, all cells were plated on YEP plate containing kanamycin, rifampicin and streptomycin, and incubated at 28°C for 2 to 3 days. Successful transformants were inoculated in 3-5 mL YEP medium containing kanamycin, rifampicin and streptomycin and incubated at 28°C for overnight before being prepared for plant transfection as described in (Sparkes *et al.*, 2006).

<u>Tobacco Growth condition:</u> Tobacco seeds were germinated in compost and grown under long day conditions (16 h light) at 21ºC. Leaves from 5-week-old plants were used for *Agrobacterium* infiltration as described in (Sparkes *et al.*, 2006). ~2 g infiltrated leaf samples were used for small scale of protein extraction and expression test.

<u>Expression scale up:</u> 2-3 leaves were infiltrated per plant at the chosen condition for target protein expression. ~30 g fresh leaf was harvested per batch of infiltration (12 plants) and snap frozen in liquid nitrogen before protein extraction.

2.2.2 Plant tissue P content Assay

2.2.2.1 P extraction

Tissue samples were homogenized with extraction buffer (10 mM Tris-HCl, 1 mM EDTA, 100 mM NaCl, 1 mM 2-mercaptoethanol and 1 mM phenylmethylsulfonyl fluoride (PMSF), pH 8.0) at a ratio of 1 mg dry sample to 100 μL of extraction buffer. A total of 100 μL of homogenized sample was mixed with 900 μL of 1% glacial acetic acid and incubated at 42°C for 30 min. 100 μl of the above mixture was transferred to a glass tube for total P assay and the rest of the solution was centrifuged at 13,000 g for 5 min and 300 μL of the supernatant aliquot was used for inorganic phosphate (Pi) content measurement.

For total P assay, 30 μL of 10% $Mg(NO_3)_2$ in 95% ethanol was added to the 100 μL aliquot. Samples were then dried and flamed to ash. After cooling, 300 μL of 0.5 M HCl was added to dissolve the samples at 65°C for 30 min. The total and inorganic P contents of the samples were then measured by the P assay method below.

2.2.2.2 P assay (Ames, 1966)

Reagents

a) Ascorbic acid, 10% (kept refrigerated up to a month).
b) 0.42% Ammonium molybdate in 1N H_2SO_4 (kept at room temperature).

P assay reagent mix: Reagent (a) and reagent (b) were mixed at a ratio of 1:6 (kept on ice up to one day).

Procedure

0.7 mL of the P assay reagent mix was added to 0.3 mL of the extracted total P or Pi solutions and incubated at 37°C for 1 h. Light absorbance was measured at 820 nm using a spectrophotometer. 0.3 mL milliQ H_2O (blank) and 0.3 mL standard Pi solutions were used to generate the calibration curve for determination of the P content in extracted solutions.

2.2.3 RNA isolation, cDNA preparation and Semi-quantitative PCR

2.2.3.1 RNA extraction

Total RNA from plant shoot and root tissue was extracted using an RNeasy Plant Mini Kit (Qiagen) according to the manufacturer's instructions. On-column DNase digestion was performed using the RNase-Free DNase Set (Qiagen, 79254) according to the instruction from the RNeasy Mini Handbook (Qiagen). The concentration of total RNA was measured from its absorbance at 260 nm using a DS-11 Spectrophotometer (DeNovix). Quality of extracted total RNA was assessed with an Agilent 2100 Bioanalyzer (Agilent Technologies) using RNA 6000 Nano Assay kit (Agilent Technologies), according to the manufacture's instruction.

2.2.3.2 RT-PCR

First-strand cDNAs were synthesized from 1 μg total RNA using Precision nanoScript TM2 Reverse Transcription kit (PrimerDesign). A mixture of oligo dT and random hexamer primers were used according to the protocol provided by the RT reaction kit (PrimerDesign). Semi-quantitative PCR was performed with KOD Hot Start DNA Polymerase (Novagen). Each 50 μL reaction contained 2 μL cDNA of a 20-μL reverse transcriptase reaction mix and 3 μL (0.6 μM) forward and reverse primers (Table 2.2). The PCR conditions started with polymerase activation by heating for 2 min at 94°C, followed by 35 cycles of 94°C for 15 sec, 60°C for 30 sec and 72°C for 45 sec. The amounts of amplified DNA were analyzed on a 1% agarose gel using 5 μL aliquots from each reaction.

2.2.3.3 Q-PCR reactions

Q-PCR reactions were carried out with a CFX96 Touch Real-Time PCR Detection System (Bio-rad) using Q-PCR reagent 'PrecisionPLUS MasterMix for the Bio-Rad iCycler premixed with SYBRgreen' (Primer Design). All Q-PCR reactions were performed in duplicate on independent RNA samples obtained from three biological replicates, using gene specific primers for potato SPX exclusive family members as well as primers for

reference gene *StL2* (Table 2.2) (Nicot *et al.*, 2005). Each 20 µL reaction contained 2 µL cDNA of a 30x diluted reverse transcriptase reaction product and 300 nM forward and reverse primers. The following conditions were used according to the manufacturer's instructions: initial heating for 2 min at 95°C, followed by 40 cycles of 95°C for 15 sec, 60°C for 1 min. Melting curves were produced for each amplicon at the end of 40 cycles by heating the reaction mix for 5s/temperature increment from 65°C to 95°C with a 0.2°C increment. Fluorescence data were collected by whole plate imaging at the end of each amplification cycle and at the end of each melting curves heating steps. Cycle threshold values and amplification efficiencies were calculated using the CFX Manager software (Bio-rad). The calculation of relative gene expression levels were described in Chapter 3 section 3.3.

Table 2.2 Potato (*Solanum tuberosum*) gene specific primers used in RT-PCR and Q-PCR reactions. (RT-F: Forward primers used for RT-PCR reactions; RT-R: Reverse primers used for RT-PCR reactions; Q-F: Forward primers used for Q-PCR reactions; Q-R: Reverse primers used for Q-PCR reactions. Q-PCR primers for reference gene *StL2* and *Stef1-a* are taken from (Nicot *et al.*, 2005)).

Gene name	Primer sequences (5' to 3')	Amplicon length
StSPX1	Q-F: CGAGCTGCAGGGTTGGTTTATC	140 bp
	Q-R: CGCTCTTCTCCTTCACACGTTC	
	RT-F: TTCCACCGGCGATAATCCAC	413 bp
	RT-R: TCGGCTGATGAAGTGCAACTTG	
StSPX2	RT/Q-F: ATGGAAGCCAACCCTCTGAG	139 bp
	RT/Q-R: TTCAGTATCTTTGCCAATCCTGTGTAG	
StSPX3	RT/Q-F: GCAAAGGCAAAAGATCGTAATGATGAG	121 bp
	RT/Q-R: TCTTCACGAGCCCAGTATAGTTG	
StSPX4	Q-F: TGGCAAAGACGGAGTTCTCAC	138 bp
	Q-R: TTAACAAGACCTGCAAAATTTAGAGAGC	
	RT-F: ACTTGCAGGACTGGTTCGTC	274 bp
	RT-R: TTAACAAGACCTGCAAAATTTAGAGAGC	
StSPX5	Q-F: TTAAGGTATTGAAAGAAAGGGTAGCTGAG	139 bp
	Q-R: ACAACTCCTGTATAGTTAAGAGCACTG	
	RT-F: TTAAGGTATTGAAAGAAAGGGTAGCTG	424 bp
	RT-R: ACTCTGAGTGCTGAATAGGTAAGTC	
StL2	Q-F: GGCGAAATGGGTCGTGTTAT	121 bp
	Q-R: CATTTCTCTCGCCGAAATCG	
Stef1-α	Q-F: ATTGGAAACGGATATGCTCCA	101 bp
	Q-R: TCCTTACCTGAACGCCTGTCA	

2.2.4 *E. coli* **DNA manipulation**

2.2.4.1 *Amplification of target genes*

Target genes were amplified using KOD Hot Start DNA Polymerase (Novagen). Polymerase chain reaction (PCR) systems were set up based on the manufacturer's instructions. Target DNA fragments were amplified by a touchdown hot start program, beginning with 2 min heating at 95°C. The main reaction contained 5 cycles of 30 s denaturation at 94°C, 30 s annealing at an initial temperature of 60°C then decreasing by 1°C per cycle and 1 min extension at 72°C; 30 cycles of 30 s denaturation at 94°C, 30 s annealing at 55°C and 1 min extension at 72°C. After a 10 min final extension at 72°C, the reaction product was held at 4°C.

2.2.4.2 *Restriction digestion*

All restriction digestion reactions were set up in the presence of the appropriate buffers according to the manufacturer's instructions (New England Biolabs). DNA fragments were ligated using T4 DNA ligase (New England BioLabs). Ligation systems and reaction conditions were set up according to the manufacturer's protocols. The amounts of DNA were adjusted to give a vector to insert molar ratio of 3:1, 1:1, 1:3 or 1:10.

2.2.4.3 *DNA analysis by agarose gel electrophoresis*

Agarose gels were prepared in 1xTAE buffer and DNA was stained with SYBR Safe DNA gel stain (Invitrogen). 1% (w/v) gels were prepared for large DNA fragments (500-10,000 bp) and 2% gels were prepared for small DNA fragments (50-500 bp). Sample DNAs were mixed with 6x sample buffer (New England Biolabs) and resolved at a constant voltage of 100V for 15-30min. Electrophoresis results were analyzed under blue light by comparing with DNA markers (New England Biolabs) electrophoresed alongside the samples.

2.2.4.4 Gel purification of DNA fragments

Desired DNA bands were excised from agarose gels using razor blades and the DNA fragments were purified and recovered using Wizard SV Gel and PCR Clean-Up System (Promega) according to the manufacturer's instructions.

2.2.4.5 Plasmid construction (Plasmid maps in Appendix 2, 3)

pEAQ-*HT*-His$_6$-AtSPX1 / pEAQ-*HT*-AtSPX1-His$_6$: For *agrobacterium* mediated tobacco expression, the cDNA encoding full-length AtSPX1 (obtained from the Arabidopsis Biological Resource Centre (ABRC), At5g20150) was sub-cloned into pEAQ-*HT* binary vector (Sainsbury *et al.*, 2009) via the *Xma*I/*Xho*I sites and *Age*I/*Xma*I sites to generate N-terminal and C-terminal His$_6$ tagged AtSPX1, respectively. Primer AtSPX1-NH_F and AtSPX1-NH_R were used for the pEAQ-HT-His$_6$-AtSPX1 construct and primer AtSPX1-CH_F and AtSPX1-CH_R were used for the pEAQ-HT-AtSPX1-His$_6$ construct.

pUC18-AtSPX1-His$_8$-GFP: For *Arabidopsis* protoplast expression, the open reading frame (ORF) of AtSPX1 and the first half of a C-terminal His$_8$ tag was amplified by a first step polymerase chain reaction (PCR) with primer AtSPX1-H8_F and AtSPX1-H8_R1, using full-length AtSPX1 cDNA (ABRC, At5g20150) as a template. The AtSPX1 ORF and full-length of C-terminal His$_8$ tag coding region was then amplified by a secondary PCR step with primer AtSPX1-H8_F and AtSPX1-H8_R2, using the first step PCR product as a template. The AtSPX1 ORF and C-terminal His$_8$ tag coding sequence was further sub-cloned into pUC18-based expression vector (provided by Prof. Alison Baker) via the *Xba*I site to generate a C-terminal His$_8$-GFP tagged plant expression construct of AtSPX1 under the control of the 35S promoter.

pET28-AtSPX1-His$_6$: The open reading frame (ORF) of AtSPX1 was amplified by PCR, using a codon optimized synthetic *AtSPX1* gene (Appendix 1) as a template. The amplified *AtSPX1* was sub-cloned into a pET28-based expression vector (Novagen) via the *Nco*I and *Xho*I sites, with a 6 histidine tag (His$_6$) at the C-terminus (plasmid provided by Prof. Stephen Baldwin).

<u>pET28-AtSPX1-S1/S2-His$_6$:</u> To generate 2 shorter versions of AtSPX1 with C-terminal His$_6$ tag, genes coding the C-terminal predicted disordered region of AtSPX1 was removed by a 2-step mutagenesis process (Edelheit *et al.*, 2009) using primer AtSPX1-S1_F/R and AtSPX1-S2_F/R.

<u>pET28-AtSPX1-SBP-His$_8$:</u> The HRV-3C cleavage site + Streptavidin Binding Peptide (SBP) + 8 histidine tag (His$_8$) coding sequences and a 3'-stop codon were amplified by PCR from a previously synthesized sequence 'ALMT1_C-terminal_domain_SBP_tag' (GenScript, provided by Prof. Stephen Baldwin) using primer SBP-His$_8$_F and SBP-His$_8$_R. PCR fragment was inserted between the *AtSPX1* ORF and the original His$_6$ region in plasmid pET28-AtSPX1-His$_6$ via the *Xho*I site, blocking the expression of the previous His$_6$.

<u>pET28-AtSPX1-His$_8$:</u> The SBP coding sequence from pET28-AtSPX1-SBP-His$_8$ was removed by a 2-step mutagenesis process (Edelheit *et al.*, 2009) using primer AtSPX1-MG_F and AtSPX1-MG_R, leaving an HRV-3C cleavage site + 8 histidine tag (His$_8$) attached to the C-terminal end of the *AtSPX1* ORF.

<u>pGEX-2T-AtSPX1:</u> Plasmid for expression of GST-AtSPX1 was generated by sub-cloning cDNA encoding full-length AtSPX1 (ABRC, At5g20150) into pGEX-2T GST Expression Vector (GE Healthcare) using GST-AtSPX1_F and GST-AtSPX1_R primers, via the *Bam*HI and *Eco*RI sites. [Constructed by MSc student Dikani Salema]

<u>pGEX-2T-AtSPX3:</u> Plasmid for expression of GST-AtSPX3 was generated by sub-cloning cDNA encoding full-length AtSPX3 (ABRC, At2g45130) into pGEX-2T GST Expression Vector (GE Healthcare) using GST-AtSPX3_F and GST-AtSPX3_R primers, via the *Bgl*II and *Eco*RI sites. [Constructed by MSc student Dikani Salema]

<u>pMAL-c2x-AtdPHR1:</u> For expression of MBP-AtdPHR1, the gene encoding a truncated version of AtdPHR1 (amino acids 208 – 362) was amplified by polymerase chain reaction (PCR) with primer AtdPHR1_F and AtdPHR1_R, using full-length AtPHR1 cDNA (ABRC, At4g28610) as a template. The PCR product of AtdPHR1 was sub-cloned into pMAL-c2x expression vector (New England BioLabs) via the *Bam*HI and *Hin*dIII sites.

Plasmids expressing SPX domains from AtPHO1;H1 were provided by Prof. Stephen Baldwin. All primers used for plasmid construction are listed in Table 2.3. All plasmid names and features are summarized in Table 2.4.

Table 2.3 Primers used in plasmid construction.

Primer name	Sequences (5' to 3')
AtSPX1-NH_F	CCCCGGGATGAAGTTTGGTAAGAGTCTCAGC
AtSPX1-NH_R	CCTCGAGCTATTTGGCTTCTTGCTCCAAC
AtSPX1-CH_F	CACCGGTATGAAGTTTGGTAAGAGTCTCAGC
AtSPX1-CH_R	CCCCGGGTTTGGCTTCTTGCTCCAAC
AtSPX1-H8_F	GCTCTAGAGCTATGAAGTTTGGTAAGAGTCTCAGCAATC
AtSPX1-H8_R1	TGATGGTGGCTGCCACGCTCGAGTTTGGCTTCTTGCTCCAAC
AtSPX1-H8_R2	GTATCTAGATGGTGATGGTGATGGTGATGGTGGCTGCCACG
AtSPX1-S1_F	GCCGTTCATCCAGAAAGTTCTGCAACTCGAGCACCACCACCACCACC
AtSPX1-S1_R	GGTGGTGGTGGTGGTGCTCGAGTTGCAGAACTTTCTGGATGAACGGC
AtSPX1-S2_F	CTGAAGCTATGCTGGACCAGCTCGAGCACCACCACCACCACCAC
AtSPX1-S2_R	GTGGTGGTGGTGGTGGTGCTCGAGCTGGTCCAGCATAGCTTCAG
SBP-His$_8$_F	ATACACTCGAGCTGGAAGTTCTGTTCCAGGGTC
SBP-His$_8$_R	GTCTCGTCGACCTATTAATGGTGATGGTGATGGTGATG
AtSPX1-MG_F	GTCCGCATATGGGCAGCGGCAGCCGTGGCAG
AtSPX1-MG_R	CTGCCACGGCTGCCGCTGCCCATATGCGGAC
GST-AtSPX1_F	GCGGGATCCATGAAGTTTGGTAAGAGTCTCAGCAATCAG
GST-AtSPX1_R	GCCGAATTCCTATTTGGCTTCTTGCTCCAACAATGG
GST-AtSPX3_F	GCGAGATCTATGAAGTTTGGAAAGAGGATTAAAGAACAGATACAAG
GST-AtSPX3_R	GCGGAATTCTCATGGAATAGGAATCGGAGAAGATAGATGAAGAG
AtdPHR1_F	GGATCCGAATTGCGACCTGTTAGCACAAC
AtdPHR1_R	AAGCTTCACCCTTTGGTAAGACCAGAGTTTTG

Table 2.4 Plasmids constructed and used in this research. (*: codon optimized sequence)

Plasmid Name	Feature
pEAQ-*HT*-His$_6$-AtSPX1	Wild-type full-length AtSPX1 for expression in tobacco leaves; N-terminal His$_6$ tagged.
pEAQ-*HT*-AtSPX1-His$_6$	Wild-type full-length AtSPX1 for expression in tobacco leaves; C-terminal His$_6$ tagged.
pUC18-AtSPX1-His$_8$-GFP	Wild-type full-length AtSPX1 for expression in *Arabidopsis* protoplast; C-terminal His$_8$-GFP tagged.
pET28-AtSPX1*-His$_6$	Codon optimized full-length AtSPX1 for expression in *E. coli* cells; C-terminal His$_6$ tagged.
pET28-AtSPX1*-S1-His$_6$	Codon optimized truncated AtSPX1 (residue 1-159) for expression in *E. coli* cells; C-terminal His$_6$ tagged.
pET28-AtSPX1*-S2-His$_6$	Codon optimized truncated AtSPX1 (residue 1-181) for expression in *E. coli* cells; C-terminal His$_6$ tagged.
pET28-AtSPX1*-SBP-His$_8$	Codon optimized full-length AtSPX1 for expression in *E. coli* cells; C-terminal SBP-His$_8$ tagged.
pET28-AtSPX1*-His$_8$	Codon optimized full-length AtSPX1 for expression in *E. coli* cells; C-terminal His$_8$ tagged.
pGEX-2T-AtSPX1	Wild-type full-length AtSPX1 for expression in *E. coli* cells; N-terminal GST tagged.
pET28-AtPHO1;H1_1-His$_6$	Wild-type N-terminal region of AtPHO1;H1 (residue 1-340) for expression in *E. coli* cells; C-terminal His$_6$ tagged.
pET28-AtPHO1;H1_2-His$_6$	Wild-type N-terminal region of AtPHO1;H1 (residue 1-384) for expression in *E. coli* cells; C-terminal His$_6$ tagged.
pL53-MBP-AtPHO1;H1_1	Wild-type N-terminal region of AtPHO1;H1 (residue 1-340) for expression in *E. coli* cells; N-terminal MBP tagged.
pL53-MBP-AtPHO1;H1_2	Wild-type N-terminal region of AtPHO1;H1 (residue 1-384) for expression in *E. coli* cells; N-terminal MBP tagged.
pGEX-2T-AtSPX3	Wild-type full-length AtSPX3 for expression in *E. coli* cells; N-terminal GST tagged.
pMAL-c2x-AtdPHR1	Wild-type truncated AtPHR1 (residue 208 – 362) for expression in *E. coli* cells; N-terminal MBP tagged.

2.2.4.6 DNA quantification

The concentration of DNA was measured from its absorbance at 260 nm using a DS-11 Spectrophotometer (DeNovix).

2.2.4.7 Competent cell transformation and plasmid preparation

One aliquot of competent cells was mixed with no more than 10% (v/v) plasmid DNA. The mixture was treated with 30 min incubation at 4°C, 30 s heat shock at 42°C and a further 2 min incubation at 4°C. 100 µL standard LB medium was added afterwards and the mixture was incubated at 37°C for 1 h. The transformation mixture was then spread on an LB-agar plate with appropriate antibiotics and incubated overnight at 37°C for the selection of successful transformants.

Selected colonies were inoculated into 5 mL standard LB medium with appropriate antibiotics, followed by overnight orbital shaking (200 rpm) at 37°C. The subsequent preparation of plasmid DNA was performed with a Wizard Plus SV Minipreps DNA Purification System (Promega) according to the manufacturer's instructions.

2.2.5 Protein expression, purification and analysis

2.2.5.1 *Protein expression test*

Protein expression test in *E. coli* cells

Screening of different expression conditions, including combinations of a range of *E. coli* host strains and culture media was first carried out on a small scale. Freshly prepared transformants containing target protein expression vectors were inoculated into 5 mL standard LB medium with appropriate antibiotics to obtain pre-cultures by overnight orbital shaking (200 rpm) at 37°C. A sample (15 µL) of the overnight pre-culture was inoculated into 3 mL of different culture media (section 2.1.11) with appropriate antibiotics in 24 deep-well plates. For auto-induction media, the plates were sealed with breathable membranes and incubated for varying length at required temperatures in a humidified plate shaker at 1300 rpm. For IPTG induction, absorption at 600 nm was measured every 30min after inoculation till OD_{600} reached 0.5-0.7. IPTG was added to a final concentration of 0.1-0.3 mM and plates were sealed and incubated as described above.

500 µL samples of the auto-induced / IPTG-induced cultures were centrifuged at 2000 g for 15 min at 4°C, and the cell pellets were frozen at -80°C for at least 1 h before 100 µL enzyme lysis solution (section 2.1.9.1) was added to each sample and mixed well. Total cell lysate was obtained by shaking the mixture at 1000 rpm for 30 min at room temperature. Soluble cell lysate was obtained from the supernatant fraction after a further centrifugation at 14,000 g for 30 min at 4°C. These cell lysate fractions were then subjected to dot blotting and/or western blotting for the detection of target protein expression.

Protein expression test in tobacco leaves

~2 g infiltrated tobacco leaves expressing target proteins were ground with pestle and mortar in liquid nitrogen. 3x volume of denaturing extraction buffer (50 mM Tris-HCl, 1% SDS, 2 mM PMSF, 10 µL/mL protease inhibitor cocktail for plant cell extracts, pH 7.5) was added to the ground samples and mixed thoroughly by vortexing. Soluble proteins

50

were obtained from the supernatant fraction of a subsequent centrifugation at 18,000 g for 10 min at 4°C. Non-soluble fractions (pellets from the centrifugation) were re-suspended in equal amount of SDS sample buffer. Quantification of soluble proteins in the crude extraction was carried out using BCA assay. For detecting the expression of target proteins, 20 μg soluble proteins and 10 μL re-suspended non-soluble fractions from each extraction was analyzed by SDS-PAGE and subsequent immunoblotting with corresponding antibodies.

2.2.5.2 *Protein purification from E. coli cells*[1]

<u>Native Immobilized Metal Affinity Chromatography (IMAC) purification of AtSPX1-His$_6$</u>

For auto-induction of AtSPX1-His6, plasmid DNA was transformed into BL21-gold (DE3) cells. Pre-culture of successful transformants (in LB medium, section 2.1.11.1) was added to expression culture with appropriate antibiotics to give a starting OD_{600} of 0.05 and then incubated at 20°C with orbital shaking at 200 rpm for up to 48 h. Cells were harvested by centrifugation at 4000 g for 30 min at 4 °C and the pellets were kept at -80°C before use.

Cell pellets were re-suspended and homogenized in ice-cold native extraction buffer (1xPBS pH 7.4, 150 mM NaCl, 10% Glycerol, 1% Triton X-100, 5 mM EDTA, 10 mM DTT and protease inhibitor cocktail) at a ratio of 6 mL buffer per gram of pellet. Cells were fully lysed by passing through a cell disruptor (Constant Systems Limited) operating at 30 kPSI at 4°C twice. Cell debris was pelleted by centrifugation at 14000 g for 45 min at 4°C. The soluble supernatant fraction was then bound to the cOmplete His-Tag Purification Resin (Roche, pre-charged with Ni^{2+}; at a ratio of 20 μL resin per mL cell lysate) with gentle mixing for at least 2 h at 4°C. The resin slurry was then packed into a column to allow the unbound fraction to pass through under gravity. The resin was subsequently washed and target protein was eluted with washing buffers (1xPBS pH 7.4, 150 mM NaCl, 10% Glycerol, 0.02% Triton X-100, 5 mM EDTA and 10 mM DTT) containing gradient concentrations of imidazole (5-500 mM) (2 column volume buffer was

---

[1] In this section, where 1xPBS is used, the NaCl concentration in buffer composition represents the concentration of additional NaCl apart from that is included in the 1xPBS buffer (section 2.1.9.1).

used for each imidazole concentration). Purification fractions of AtSPX1-His$_6$ were then subjected to SDS-PAGE analysis and western blotting with anti-His$_6$ antibody.

Denaturing IMAC purification of AtSPX1-SBP-His$_8$/ AtSPX1-His$_8$ from inclusion bodies

Plasmid DNA expressing AtSPX1-SBP-His$_8$ and AtSPX1-His$_8$ were transformed into BL21-gold (DE3) expression strain for auto-induction in SB medium as described above. Denaturation and refolding of insoluble protein from inclusion bodies was performed according to the method described in Rudolph and Lilie, 1996. Briefly, cell pellets were re-suspended and homogenized in homogenization buffer (1xPBS pH 8.0, 150 mM NaCl, 10% glycerol, 1% Triton X-100, 5 mM EDTA, 10 mM DTT) containing protease inhibitor cocktail (6 mL buffer per gram cells), and lysed using the same method as described above. After cell disruption and centrifugation, the pellet was washed by full re-suspension in inclusion body washing buffers (same volume as homogenization buffer) in the following order: washing buffer I (1xPBS pH 8.0, 150 mM NaCl, 10% glycerol, 2% Triton X-100, 5 mM EDTA, 10 mM DTT), washing buffer II (1xPBS pH 8.0, 863 mM NaCl, 10% glycerol, 1% Triton X-100, 5 mM EDTA, 10 mM DTT), washing buffer III (1xPBS pH 8.0, 150 mM NaCl, 10% glycerol, 5 mM EDTA, 10 mM DTT). Following each re-suspension the inclusion bodies were pelleted by centrifugation at 14000 g for 20 min at 4°C and finally re-suspended in the same amount of solubilizing buffer (1xPBS pH 8.0, 150 mM NaCl, 10% glycerol, 5 mM EDTA, 10 mM DTT, 6 M guanidine hydrochloride (GndCl)). Protein solubilization was carried out by 30 min incubation at 37°C with gentle shaking before remaining cell debris was pelleted by a final centrifugation at 14000 g for 20 min at 4°C. Denatured soluble proteins were purified via IMAC using the same resin described above with overnight binding. The resin was subsequently washed with solubilizing buffer containing low concentrations of imidazole (5-15 mM), and target proteins were eluted from the resin with solubilizing buffer containing a higher concentration of imidazole (300 mM). IMAC elution fractions were dialyzed overnight against dialyzing buffer (1xPBS pH 8.0, 150 mM NaCl, 10% Glycerol, 5 mM EDTA, 10 mM DTT, 400 mM L-Arginine) to remove GndCl and imidazole and soluble refolded proteins was obtained by a centrifugation at 100,000 g for 30 min at 4°C.

Purification fractions of AtSPX1-SBP-His$_8$ and AtSPX1-His$_8$ that did not contain 6 M GndCl were subjected to SDS-PAGE analysis and western blotting with anti-His$_6$ antibody. Purification fractions that contained 6 M GndCl were analyzed by dot blotting with anti-His$_6$ antibody.

Native purification of GST-AtSPX1/GST

Expression of GST-AtSPX1 and GST was induced in BL21 (DE3) cells with 0.1 mM IPTG at OD$_{600}$ ~ 0.5. Induced cells containing GST-AtSPX1 or GST were grown at 18°C overnight or at 37°C for 2 h, respectively. Cells were harvested as described above (section: IMAC purification of AtSPX1-His$_6$). Cell pellets were re-suspended and homogenized in ice-cold 1xPBS (pH 7.4) buffer containing protease inhibitor cocktail, at a ratio of 6 mL buffer per gram of pellet, and lysed as described above. Cell debris was pelleted by centrifugation at 12000 g for 20 min at 4°C. The soluble GST-AtSPX1 and GST proteins were affinity purified using Glutathione Sepharose 4 Fast Flow resin (GE Healthcare) according to the manufacturer's instructions.

The affinity purified GST-AtSPX1 was dialyzed into ion-exchange chromatography starting buffer (10 mM HEPES pH 7.5[2], 50 mM NaCl) for overnight. Dialyzed proteins were further purified via ion exchange chromatography, using a HiTrap Q-HP 1 mL column (GE Healthcare) with an ÄKTA Explorer (GE Healthcare). The column was washed with a 50 mM to 1 M NaCl gradient in 30 column volumes in 10 mM HEPES buffer (pH 7.5). 1 mL ion-exchange chromatography fractions were collected. Purification fractions of GST-AtSPX1 and GST were then subjected to SDS-PAGE analysis and western blotting with anti-GST antibody.

Native purification of MBP-AtdPHR1

Expression of MBP-AtdPHR1 was carried out in BL21-gold (DE3) cell strain according to the 'pMAL protein fusion and purification system' instruction manual (New England

---

[2] Ion-exchange chromatography buffer of pH 8 was also used with the attempt to increase GST-AtSPX1 protein stability. Details on the buffer components are included in the result section in Chapter 5.

BioLabs). Cells were harvested and lysed as described above. MBP-AtdPHR1 protein was affinity purified at 4°C using Amylose Resin (New England BioLabs) according to the manufacturer's instructions. The affinity purified MBP-AtdPHR1 was dialyzed into ion-exchange chromatography starting buffer (10 mM HEPES pH 7.5, 50 mM NaCl) overnight. Dialyzed proteins were further purified via ion exchange chromatography, using a HiTrap Q-HP 1 mL column (GE Healthcare) with an ÄKTA Explorer (GE Healthcare). The column was washed with a 50 mM to 1 M NaCl gradient in 30 column volumes in 10 mM HEPES buffer (pH 7.5). 1 mL ion-exchange chromatography fractions were collected. Purification fractions of MBP-AtdPHR1 were then subjected to SDS-PAGE analysis.

Size exclusion chromatography on ion exchange chromatography purified MBP-AtdPHR1 was performed on a Superdex 200 10/300 GL column (GE Healthcare) with same buffer composition for SPR assays (10 mM HEPES pH 7.5, 200 mM NaCl). The void volume of the column was determined using Blue Dextran 2000 and the partition coefficient K_{av} for individual proteins was calculated as follows: $K_{av} = (V_r-V_o)/(V_c-V_o)$ according to the Superdex 200 HR 10/30 instruction (GE Healthcare). A calibration curve of protein molecular weight was generated by plotting Log Mw against K_{av} values for standard proteins Ovalbumin (44 kDa), Conalbumin (75 kDa), Ferritin (440 kDa), Aldolase (158 kDa) and Thyroglobulin (669 kDa).

2.2.5.3 *Protein purification from tobacco leaves*

<u>Native IMAC purification of AtSPX1-His$_6$</u>

Frozen leaf samples were ground with pestle and mortar in liquid nitrogen before being added into 1.5x volume of ice cold native extraction buffer (50 mM Tris-HCl pH 6.8, 150 mM NaCl, 0.1% Triton X-100, 1 mM PMSF and plant protease inhibitor cocktail) and mixed thoroughly by vortexing. Total cell lysate was centrifuged at 4°C, 18,000 g for 10 min to remove plant debris. The soluble fraction was obtained by filtering the supernatant fraction through 0.45 μM filters. The filtered soluble proteins were bound to Ni-NTA resin (Qiagen) at 4°C for 2 h, before loading on to a gravity flow column. The column was washed with 5 column volume of washing buffer (50 mM Tris-HCl pH 6.8, 150 mM NaCl)

including 5 mM imidazole and then 30 mM imidazole. Target protein was eluted with 1 column volume of eluting buffer (50 mM Tris-HCl pH 6.8, 150 mM NaCl, 300 mM imidazole). Purification fractions of AtSPX1-His$_6$ were then subjected to SDS-PAGE analysis and western blotting with anti-His$_6$ antibody.

2.2.5.4 SDS-PAGE and Immunoblotting

Protein purification fractions and pull-down fractions were analyzed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Samples were suspended in SDS loading buffer, incubated at 95°C for 5 min and run on 12% polyacrylamide gels (section 2.1.9.2). Protein bands were stained by Coomassie blue staining (section 2.1.9.4) before imaging.

For western blotting, SDS gels were submerged in blotting buffer for at least 15 min before proteins were transferred to an Amersham Nitrocellulose Western blotting membrane (GE Healthcare) using a Trans-Blot Turbo Transfer System (Bio-rad) for 20 min at 25V. Membranes were blocked in blocking buffer (section 2.1.9.3) for 2 h at room temperature. The membranes were probed with corresponding antibodies for at least 2 h before being washed with TBST buffer for three times (15 min / wash). Membranes were then incubated with a freshly-prepared mixture of SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher Scientific) for about 1 min and chemiluminescence was detected and quantified using a GeneGnome Detection system and GeneTools software, respectively (Syngene Bio Imaging).

For dot blotting, samples were suspended in SDS loading buffer, incubated at 95°C for 5 minutes before 2-3 μL was dotted onto nitrocellulose membranes. When samples were dry, the nitrocellulose membranes were blocked, probed, washed and developed as described above for western blotting.

2.2.5.5 Protein concentration determination

BCA

The concentration of crude protein extracts was measured using the Pierce BCA Protein Assay Kit (Thermo Fisher) according to the manufacturer's instructions. Standard bovine serum albumin (BSA) protein (Sigma) was diluted in a 2-fold series to generate the calibration curve. 10 µL of samples and standards were added to a 96-well plate in duplicate and then 200 µL assay reagent (4% $CuSO_4$: BCA reagent A (Pierce) = 1:50) was added to each well. The plate was incubated at 37 ºC for 30 min and then the absorbance was measured at 570 nm using a plate reader (ICN Biomedical Titertek).

A280

The concentration of purified proteins was determined by measuring the absorption at 280 nm, using a Biowave II spectrometer (BioChrom Ltd.) or a DS-11 Spectrophotometer (DeNovix). Protein concentrations were calculated following the Beer-Lambert Law:

$$A=\varepsilon cl$$

[A: absorption at 280 nm; ε: molar extinction coefficient; c: concentration; l: length of light path]

The predicted molar extinction coefficients of the proteins were calculated using the ExPASy website (http://www.expasy.org).

2.2.6   Protein characterization

2.2.6.1   *Pull-down assay*

To qualitatively detect the interaction between GST-AtSPX1 and MBP-AtdPHR1, MBP-AtdPHR1 was expressed and purified as described above except the lysis buffer was either with Pi (1xPBS pH 7.4, 63 mM NaCl, 1 mM EDTA, 1 mM DTT) or without Pi (20 mM Tris pH 7.5, 200 mM NaCl, 1 mM EDTA, 1 mM DTT) in the presence of cOmplete EDTA-free Protease Inhibitor Cocktail (Roche). MBP-tagged constructs in the cell lysates were immobilized to amylose affinity matrix for 2 h at 4°C. MBP-AtdPHR1 bound resin was washed with 8 column volume of the same lysis buffers and incubated with affinity purified GST-AtSPX1 for 2 h at 4°C. Any unbound proteins were removed by washing with 6 column volume of the same lysis buffers and bound proteins were eluted with 3 column volume of same lysis buffers containing 10 mM maltose.

In the control pull-down experiment, affinity purified GST-AtSPX1 was replaced with affinity purified GST protein. Samples of unbound flow-through and elution fractions were suspended in SDS gel loading buffer and analyzed on 12% SDS-PAGE gels. The presence of GST-AtSPX1/GST proteins in the fractions was detected by western blotting with anti-GST antibody.

2.2.6.2   *Surface plasmon resonance (SPR)*

DNA probe preparation

P1BS probes containing 1x P1BS motif and 2x P1BS motifs were taken from the promoter regions of AtSPX3 (-192 bp to -139 bp) and AtSPX1 (-136 bp to -83 bp), respectively (Table 2.5). Single forward strand DNA molecules were synthesized and biotinylated (Sigma-Aldrich) on the 5' end (Table 2.5). Reverse complementary strands were synthesized without biotinylation. The control DNA probe with met-box sequences does not contain P1BS motifs.

Table 2.5 DNA probes containing P1BS motifs used for SPR assays. P1BS motifs are underlined.

Probe	Sequence	length
1xP1BS	ACACTTCGTCACGCTAAAGCTAA<u>GCATATCC</u>GCTTTCATATT CCTTTACACAAC	54 bp
2xP1BS	CAGAGAAAAAA<u>GGATATTC</u>TAATTAGAAACCTTAA<u>GAATAT TC</u>TTTTTAATCCC	54 bp
Control	CCGGCAGGAGACGTCTAGACGTCTCCGGCAGG	32 bp

DNA oligos were dissolved to 100 μM in TE buffer according to manufacturer's instructions (Sigma-Aldrich). Single strand biotin-labeled and non-biotin-labeled DNA oligos were diluted in TM buffer (10 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$) to a final concentration of 10 μM and 11 μM, respectively, before annealing at 95°C for 1 min and cooling slowly down to room temperature. Annealed DNA probes were kept at -20°C before use.

SPR assays

Surface plasmon resonance (SPR) experiments were performed on a Biacore 3000 instrument (GE Healthcare). Biotinylated DNA probes at 10 nM were immobilized on Streptavidin (SA) sensor chips (GE Healthcare) with contact times of 6-8 min and a flow rate of 5 μL/min, to give approximately 500 RU of immobilized DNA. The reference flow cell was un-derivatized. All ligand immobilization was done in HEPES-buffered saline consisted of 10 mM HEPES pH 7.5, 200 mM NaCl, 0.01% Surfactant P-20. Analyte measurements were carried out at 25°C and a flow rate of 40 μL/min, using the same buffer with or without additional Pi/InsP6. For MBP-AtdPHR1 – DNA binding assays, 120 μL IEX purified MBP-AtdPHR1 was injected in a 2-fold ascending concentration, commencing at 0.3125 nM and ending at 80 nM. For sequential injection experiments, the 'Coinject' command (Biacore 3000 Instrument Handbook, GE Healthcare) was used to inject 100 μL of the first component immediately followed by 120 μL of the second component. For pre-mixing injection experiments, 120 μL pre-mixed proteins were injected over the chip surface using the 'Inject' command. The chip surface was

regenerated between protein injections with a 40 µL 0.05% SDS injection. The binding data was analyzed using the BIAevaluation 3.1 software (GE Healthcare).

2.2.6.3 *Microscale Thermophoresis (MST)*

<u>MST technique outline:</u> MST uses the directed movement of molecules in a microscopic temperature gradient to measure the biomolecular interactions. During the thermophoresis assays, a local temp difference is generated inside the sample containing capillary by an infrared laser, leading to the change of molecule concentration in the heated area. Thermophoresis of fluorescent labeled molecules in the heated spot is monitored during the titration of an unlabeled ligand, and the change in thermophoresis is used to quantify the binding.

<u>Labelling of GST-AtSPX1 and GST proteins:</u> Affinity purified GST-AtSPX1 and GST were dialyzed 2 times against 500x volume of labelling buffer (50 mM Tris-HCl pH 8.0, 1 mM TCEP (tris (2-carboxyethyl) phosphine) to remove DTT and Glutathione. Dialyzed GST-AtSPX1 and GST proteins were labelled on the thioyl groups that are located in GST protein, using Alexa Fluor 488 C_5 Maleimide (Thermo Fisher, A10254) that was dissolved in dimethyl sulfoxide (DMSO) (Fluka, 41647). The labelling dye and proteins were mixed at a 1:1 molar ratio and the labelling reaction was carried out on ice for 1 h before unincorporated dye was removed with a NAP-5 column (GE Healthcare). NAP-5 column was prepared and used according to the manufacturer's instruction and labelled GST-AtSPX1 and GST proteins were eluted in the same buffer used in SPR experiments (10 mM HEPES pH 7.5, 200 mM NaCl). Eluted proteins were collected in 200 µL fractions and the fraction with the most intense color from each protein (by visual observation) was used for the following MST experiments.

<u>MST experiments:</u> MST experiments were carried out using Monolith NT.115 instrument (Nano Temper technologies). Labelled GST-AtSPX1 and GST proteins were diluted to give fluorescence intensity <1500 in the final reaction when using 50% to 80% LED (green) power. 0.2 M Pi solution (prepared by mixing 0.2 M Na_2HPO_4 and NaH_2PO_4 to achieve a pH 7.4 solution) was used as the unlabeled binding partner for the binding assays. A 16 step 2x fold dilution series was prepared for the unlabeled binding partner

and MST experiments were performed according to the manufacturer's instructions (User Manual for the Monolith NT.115, Nano Temper technologies).

The MST data was collected and analyzed using the NT Control Software (Nano Temper technologies). All MST laser powers (20%, 40% and 80%) were tested according to the manufacturer's instructions (User Manual for the Monolith NT.115, Nano Temper technologies).

2.2.6.4   Optim analysis of protein unfolding and aggregation

Stepped thermo-unfolding and aggregation test on GST-AtSPX1 protein was carried out using Avacta Optim 1000 instrument (Isogen Life Science). Affinity purified GST-AtSPX1 protein was mixed with a range of different screened buffers (Table 2.6) and ~1 µL mixed sample was loaded to the micro-cuvette array (MCA) before heating from 15°C to 90°C with a 1°C step. The temperature at each step was held for 30 s and protein aggregation was examined by static light scattering (SLS) at 266 nm. Additional instrument settings of Optim 1000 included 1000 ms exposure, 380 nm center wavelength and 100 µm slit width. The SLS data was collected and analyzed using the Optim software package (Isogen Life Science).

Table 2.6 Optim 1000 micro-cuvette array (MCA) screened conditions

Condition	Reagent	Volume ratio (protein: reagent)
pH range	pH 3.0 – 3.5: Citric acid / sodium citrate	9:1
	pH 4.0 – 5.0: Acetic acid / sodium acetate	
	pH 5.5 – 6.5: MES / NaOH	
	pH 7.0 – 8.0: HEPES / NaOH	
	pH 8.5: BICINE / NaOH	
	pH 9.0 – 9.5: CHES / NaOH	
Glycerol concentration range	0% Glycerol (MilliQ H_2O)	1:1
	5% Glycerol	
	10% Glycerol	
	20% Glycerol	
	30% Glycerol	
	40% Glycerol	
	50% Glycerol	
	60% Glycerol	
NaCl concentration range	0 mM NaCl (MilliQ H_2O)	1:1
	100 mM NaCl	
	300 mM NaCl	
	500 mM NaCl	
	1000 mM NaCl	
	1500 mM NaCl	
	2000 mM NaCl	
	3000 mM NaCl	
Additives	Ammonium sulphate (VWR)	9:1
	Potassium acetate (BDH)	
	Potassium glutamate (Sigma)	
	Potassium chloride (Fisher)	
	Sodium sulphate (Sigma)	
	Sodium chloride (Acros)	
	Lithium chloride	
	Magnesium chloride	
	Magnesium nitrate (Sigma)	
	Calcium chloride (Sigma)	
	Lithium nitrate (Acros)	
	Sodium thiocyanate (Alfa Aesa)	
	Lithium perchlorate (Sigma)	
	Arginine (Sigma)	

Chapter 3   Physiological responses and expression of SPX exclusive genes in potato (*Solanum tuberosum*) under phosphate starvation

3.1   Introduction

Previous research on SPX domain-containing proteins has established the importance of this SPX protein domain in plant Pi regulation pathways (Secco *et al.*, 2012a). While the structural studies provide insights into the molecular basis for how this conserved SPX domain participates in Pi signalling, information on gene expression profiles is also urgently needed. The latter helps to gain a better understanding on how the SPX domain-containing proteins are expressed in plants, especially crop plants Pi starvation response cascades on a cellular level. Potato is a globally important agricultural crop grown on all 5 continents, and production in the developing world has increased rapidly in the past decade. Potato has a particularly high requirement for P fertilizer (Holmes, 2012). The availability of genome sequence from potato (Xu *et al.*, 2011), gives the opportunity to start to investigate molecular responses to P deficiency and in particular the role of SPX domain proteins.

In this chapter a hydroponic system in the greenhouse was used to study the Pi starvation responses of potato (*Solanum tuberosum*) under tightly controlled nutrient conditions. The spatial and temporal expression patterns of SPX exclusive family corresponding genes, and their response to environmental Pi levels was also investigated using quantitative PCR. These gene expression results, correlated with measurements of Pi content in tissue samples during various growth stages provides useful information on potato Pi stress responses and the involvement of SPX domain-containing proteins during this process. As genetic manipulation on *OsPHF1* promoter sequence has shown to significantly enhance Pi uptake in rice (Ruan *et al.*, 2015), potential gene modification candidates could also be identified from this study to develop potato plants with improved phosphate use efficiency (PUE).

3.2 Results

3.2.1 Physiological adaptation of potato plants during Pi starvation response

In order to closely control the nutrient input on the experimental condition as well as conveniently examine the plant root system, a small scale hydroponic system was used by adapting the conditions used in previous research on potato plants (Hammond *et al.*, 2011). Potato tubers were germinated in perlite with deionized water before 6 plantlets of 5-10 cm height were transferred into the hydroponic system and supplied with 300 μM Pi. After the plants were established in a hydroponic system (10 days), Pi deprivation conditions were introduced to half of the experimental plants by reducing the Pi concentration from 300 μM to 10 μM. After the different Pi treatment conditions were applied, the pH of the hydroponic solution was adjusted daily while the complete solution was changed twice a week to maintain corresponding Pi levels. Physiological changes of hydroponically grown potato plants were observed over a total experimental course of 28 days.

Morphological adaptation of potato plants under Pi deprivation condition

Root architecture alteration is commonly observed upon Pi starvation treatment (Sato and Miura, 2011). In order to observe the alterations of the potato root system, plant root tissues were examined 3 weeks after transferred to low Pi conditions. Plants maintained under high Pi (300 μM) treatment showed prominent primary root development with few root hairs in the elongation zone (Figure 3.1 A). In contrast, an increased growth of secondary roots and increase in root hair length and density were observed from the plants under low Pi (10 μM Pi) treatment (Figure 3.1 B), which is consistent with the previous observations of *Arabidopsis* root architecture reprogramming during Pi limited conditions (Yuan and Liu, 2008).
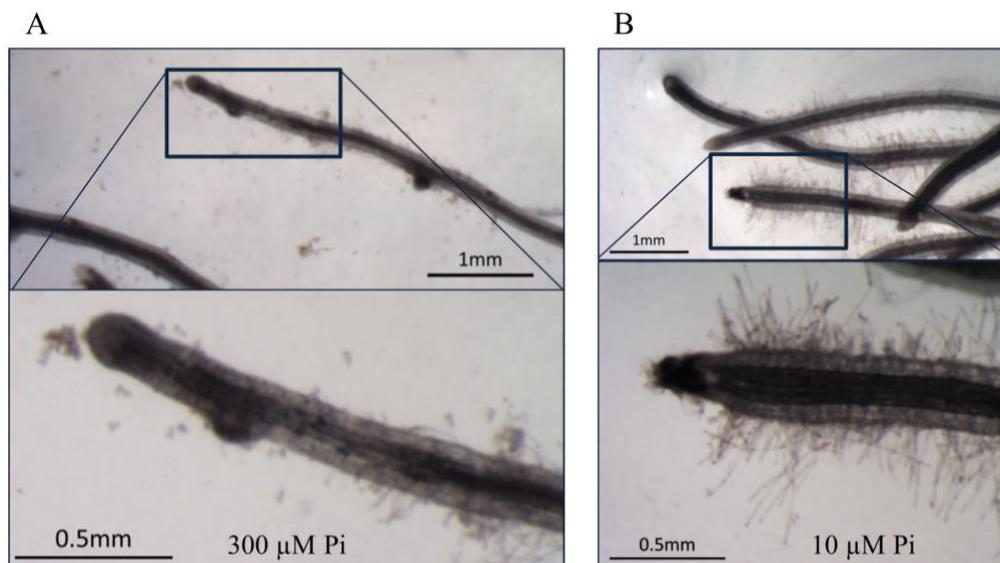
Figure 3.1 Root architecture changes of potato plants under altered Pi conditions.

Root architecture changes of hydroponically grown potato plants were observed 22 days after altering Pi conditions. A. Root tissue from plants provided with 300 µM Pi; B. Root tissue from plants provided with 10 µM Pi.
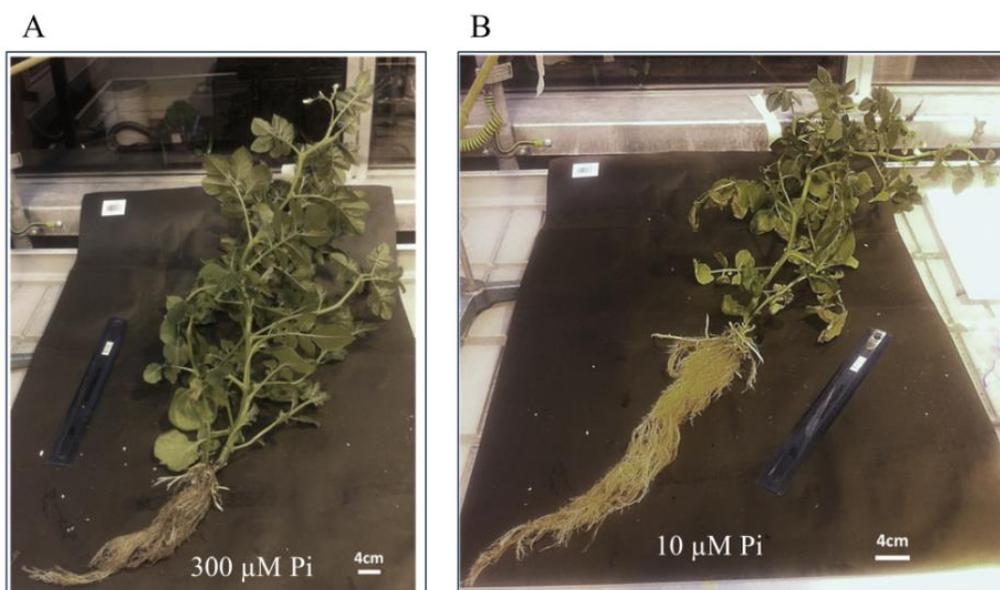
Figure 3.2 Morphological changes of potato plants under altered Pi conditions.

Whole plant morphological changes of hydroponically grown potatoes were observed 28 days after altering Pi conditions. A. Plant provided with 300 µM Pi; B. Plant provided with 10 µM Pi.

Pi starvation treatment also resulted in reduced shoot growth (Figure 3.2 A and B). This, together with the increased root growth from the low Pi treatment plants resulted in a strongly altered shoot to root ratio compared to the high Pi treatment plants (Figure 3.2).

Pi content change in potato tissues under Pi deprivation condition

To examine the effect of Pi deprivation on plant phosphorus content, tissues from hydroponically grown potatoes were harvested from both high and low Pi treatment groups on the 1st, 7th, 14th, 21st and 28th day. Both the total P and the inorganic (Pi) concentration in both leaf and root tissues were measured (Ames, 1966) (Figure 3.3). For samples grown with sufficient Pi nutrient in solution, the P assay showed a total P content around 0.8% of tissue dry weight was maintained in potato leaf samples. However, when Pi is depleted from the growing solution, the total P content dropped over a period of 14 days to 0.2% of dry weight in leaf tissues (Figure 3.3 upper panel) where it was maintained for the remainder of the experiment. P assay on potato root samples showed similar total P content changes from ~1% dry weight to between 0.2% - 0.3% dry weight after prolonged Pi starvation (Figure 3.3 lower panel). Inorganic P content in leaf and root tissues is also found to follow the same trend of decreasing under low Pi condition. As the Pi content is found to take a large proportion of the total plant P content (Figure 3.3), the reduction of Pi stores is responsible for most of the reduction in total P.

Monitoring the P content at different sampling points revealed that P content stays unchanged (P>0.1, n=3) within the first day of Pi withdrawal (Figure 3.3). Both total and inorganic P content in the shoot and root tissues was decreased significantly within 7 days of Pi starvation treatment (P<0.01, n=3) and maintained at the lower level towards the end of the experimental period (Figure 3.3). It is of notice that a more dramatic P content change in root tissues was observed within the first week of Pi starvation, whereas leaf P content decreased gradually and stabilized at a lower level over 2 weeks of low Pi treatment. This suggests the root tissue P content might be more sensitive to environmental Pi availability. A slight decrease in P content was observed from the high Pi treatment leaf and root during the last week of experiments (Figure 3.3), suggesting possible growth restriction at a later stage of plant development due to the limited growing

space from the hydroponic system. Nevertheless, the above physiological changes from potato plants have established the hydroponic system used is capable of reproducibly producing Pi starvation conditions. The gene expression changes of potato SPX domain-containing proteins under Pi limited conditions were thus investigated using this system.
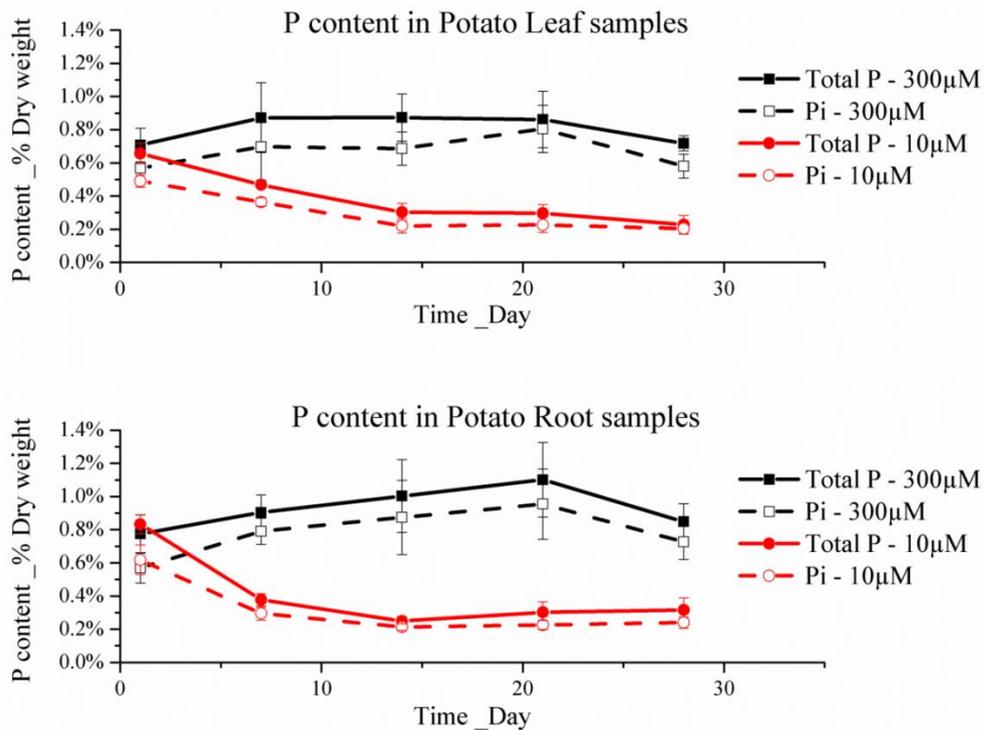
Figure 3.3 Total and inorganic P content in potato leaf and root tissue samples.

Total (Total P) and inorganic P (Pi) content in hydroponically grown potato leaf (upper panel) and root (lower panel) tissues was measured using method described in (Ames, 1966). Values are expressed as means ± SD as percentage of tissue dry weight.

3.2.2 Identification of genes encoding potato SPX domain proteins and characterization of SPX exclusive gene expression during Pi starvation response

Identification and classification of SPX domain-containing proteins

In order to understand how the conserved SPX domain is involved in Pi starvation responses in potato, SPX domain-containing proteins were identified using bioinformatics. Searching of the Phytozome website (http://www.phytozome.net) with the key word 'SPX' and BLAST searches using the AtPHO1;H1-SPX region sequence revealed the presence of 14 probable SPX domain-containing proteins in the potato (*Solanum tuberosum*) (Table 3.1). Alignments of these potato SPX domain-containing protein sequences with those from rice and *Arabidopsis* suggests that all are correctly predicted full length sequences except for StPHO1;1 and StPHO1;3, suggesting possible mis-prediction of intron splice sites for these two proteins (Figure 3.4, Appendix 4, 5). Full length sequences of these two proteins were manually corrected and 5 SPX-exclusive proteins, 5 SPX-EXS proteins, 2 SPX-MFS proteins and 2 SPX-RING proteins were subsequently classified into each of the plant SPX domain-containing protein families, using the corrected sequences (Figure 3.4 B, Appendix 5).

Sequence alignment result shows the five identified potato SPX-exclusive proteins share conserved amino acid sequences with those from other species, especially within the 3 sub-domains, suggesting that they are genuine SPX exclusive proteins, rather than domains attached to other proteins that resulted from gene model mis-prediction. All the conserved amino acid residues from the phosphate binding cluster (PBC) and lysine surface cluster (KSC) of the SPX domain (Wild *et al.*, 2016) are also found in the five potato SPX-exclusive proteins (Figure 3.4 A), further indicating these identified potato homologues might also share the key structural features with the SPX domains from other proteins. Although this research focuses primarily on the SPX exclusive family members, protein sequences alignment with already identified SPX proteins in rice and *Arabidopsis* showed the presence of all highly conserved PBC residues and most KSC residues in potato homologues in the four SPX domain-containing protein families (Figure 3.4 A, Appendix 5). This shows the pervasiveness of this conserved SPX domain across different

species, and also indicates the indispensability of highly conserved residues for the function of SPX domain-containing proteins.

Table 3.1 Potato (*Solanum tuberosum*) SPX domain-containing proteins identified from Phytozome website (http://www.phytozome.net)

Proteins with manually corrected sequences were marked with *.

SPX domain containing protein family	Identified *Solanum tuberosum* member	Phytozome protein ID	Corresponding gene locus name	Corresponding transcript ID
SPX exclusive	StSPX1	PGSC0003DMP400043620	PGSC0003DMG400025135	PGSC0003DMT400064708
	StSPX2	PGSC0003DMP400045130	PGSC0003DMG400026017	PGSC0003DMT400066925
	StSPX3	PGSC0003DMP400005212	PGSC0003DMG400002890	PGSC0003DMT400007496
	StSPX4	PGSC0003DMP400002400	PGSC0003DMG400001340	PGSC0003DMT400003384
	StSPX5	PGSC0003DMP400027779	PGSC0003DMG400015850	PGSC0003DMT400040971
SPX-EXS	StPHO1;1*	PGSC0003DMP400002589 PGSC0003DMP400002590	PGSC0003DMG400001430	PGSC0003DMT400003628 PGSC0003DMT400003629
	StPHO1;2	PGSC0003DMP400029975	PGSC0003DMG400017163	PGSC0003DMT400044207
	StPHO1;3*	PGSC0003DMP400043610	PGSC0003DMG400025127	PGSC0003DMT400064691
	StPHO1;4	PGSC0003DMP400049401	PGSC0003DMG400028396	PGSC0003DMT400073008
	StPHO1;5	PGSC0003DMP400063376	PGSC0003DMG400041272	PGSC0003DMT400091701
SPX-MFS	StSPX-MFS1	PGSC0003DMP400039445	PGSC0003DMG402022752	PGSC0003DMT400058570
	StSPX-MFS2	PGSC0003DMP400022159	PGSC0003DMG400012523	PGSC0003DMT400032620
SPX-RING	StNLA1	PGSC0003DMP400039648	PGSC0003DMG400022866	PGSC0003DMT400058855
	StNLA2	PGSC0003DMP400020192	PGSC0003DMG400011399	PGSC0003DMT400029648

A


```
              |------------ subdomain 1 ------------|                                                                      |----- subdomain 2 --
         10        20        30        40        50        60        70        80        90       100       110
    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OsSPX5  MKFGKRLKRQIEESLPEWRDHFLNYKELKRRLNAVSSPD-------------------------------------------------PAAEARFLALLHAEVDKFNAFFLEQEED
OsSPX6  MKFGKLLKRQIEQSLPEWRDKFVSYKELKRIVASISGS--------------------------------------------------PADEAAFVAALAADIDKIDSFFLEQEEE
OsSPX3  MKFGKRLKKQVEESLPEWRDKFLAYKRLKLVRLVSSSS------------------GDVGGG------------GGGEAEFVRLLDGEVDRINAFFLEQEEE
StSPX2  MKFGKRLKQQIQETLPGWGDKFLSYKELKLVKLISSYA-----------------PQMMSGSL---------EYGKAEAEFVYLLNNEIDKFNAFFMEQEED
AtSPX3  MKFGKRIKEQIQESLPEWRDFLRYKELKN---LISSPA--------------------P---------------------VESIFVGLLNAEIDKFNAFFVEQEED
OsSPX1  MKFGKSLSSQIVETLPEWRDKFLSYKDLKRLKLIGGG----GGGGE--ERQAKRARVAADGGEEEAAAAA-----------MTPEEAGFMRLLEAELDKFNSFFVEKEEE
OsSPX2  MKFGKSLSSQIVEMQPEWRDNFLSYKDLKRLNLISGG----AAG--ERASKRRRVGGATAVTVTAAAAG---------GMTLEQAGFVGLLDAELDKFNFFFLEKEEE
StSPX3  MKFGKSLSNQIEETLPEWRDKFLSYKDLKKLKLIEPKRSLSSGDGDDVRASKRMKMAAGGFDKAPEKEP-----------MTDAEVEFVNLLEDELEKFNSFFVEKEEE
AtSPX1  MKFGKSLSNQIEQTLPEWQDKFLSYKELKRLKLIGSK----TAD----RPVKRLRLDE--FSVGISKE------------EINFIQLLEDELEKFNNFFVEKEEE
AtSPX2  MKFGKSLSNQIEETLPEWRDKFLSYKELKKKLKLMEPR----SVEN--RPNKRSRSDSNSVDTDPTVG-----------MTKEELDFISLLEDELEKFNSFFVEQEEE
StSPX5  MKFWKILKSHIEETLPEWRDKFLCYKDLKELKLIYPQD---------ARPMKRQRLNNDEVTKEVN------------------DFVKLLAEEIDKFNDFFVDKEED
StSPX1  MKFGKEFSIHLEETLPEWRDKYLCYKPLKLLIKHIP------------SSTG-DNPPPLPPLP-----------PATAAAATELQGWFIRILNEELDKFNDFYVDKEED
StSPX4  MKFGKEFTTHLEETLPEWRDKYLCYKPLKLLKNLP------------PTDGGDNPPP-------------------HLQDWFVRILNEELDKFNDFYVDKEEE
AtSPX4  MKFGKEFRTHLEETLPEWRDKFLCYKPLKLLKYYPYYS----ADFGPANSDHNDSRPVFADTTNISSAADDGGVVPGVRPSEDLQGSFVRILNDELEKFNDFYVDKEED
OsSPX4  MKFGKDFRSHLEETPAWRDKYLAYKSLKKLIKNLP-------------PDG-DPPPVAAAAEVPAGD-------GDGDGGIALGNWFARVLDMELQKLNDFYIEREEW
C-C     ***  *  :    ::  :    *  * *:::  **  **.                                              *    *   ::::::*:::::**
```

```
              |------------                                                                  |------------ subdomain 3 ------------|
         120       130       140       150       160       170       180       190       200       210       220
    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OsSPX5  FVIRQR----------ELQERIQSSS-------------SAAAEMEGRVRREVVDLHGEMVLLLNYSSINYTGLAKILKKYDKRTGGVLRLPVIAGVL-RQPFYATDLLSS
OsSPX6  FVIRHRARTPIRFNSFELQEAIKKAA-----------EAAAEVAG-IRREIVDFHGEMVLLLSYSSINYIGVGKILKKHDKRTGGALAAPVAEAVRERRHFFKTETVSR
OsSPX3  FVIRQR----------ELQETVEKVAGGGGGG----RRPAAAEMRR-VRREIVDFHGEMVLLNYSAVNYTGLAKILKKYDKRTGRLLRLPFIEKVL-RQPFFTTELISR
StSPX2  FIIRHK----------ELQQRIGRVIDTWGPNGSQPSETDCKEELAKIRKDIVDFHGEMVLLINYSNINYTGLAKILKKYDKRTGGLLRSPYIQKVL-QQPFFTTDLISK
AtSPX3  FIIHHK----------ELQYRIQRLVEKCGHNDEMS-----RENISEIRKDIVNFHGEMVLLVNYSNINYTGLAKILKKYDKRTGGLRSPFIQKVL-HQPFFKTDLVSR
OsSPX1  YIIRQK----------ELQDRVARAAG-----------RESKEELMRVRKEIVDFHGEMVLLENYSALNYTGLVKILKKYDKRTGALIRLPFIQKVL-QQPFFTTDLLYK
OsSPX2  YVIRQK----------ELRERKMASA-------------EEVMRVRKEIVDLHGEMVLLENYSALNYTGLVKILKKYDKRTGSMIRLPFVQKVL-QQPFFTTDLLYK
StSPX3  YIIRLK----------ELQDRVAKAK-----------DRNDEILKIRREIVDFHGEMVLLENYSALNYTGLVKILKKYDKRTGALLRLPFIQKVL-QQPFFTTDLLYK
AtSPX1  YIIRLK----------EFRDRIAKAK-----------DSMEKMIKIRKEIVDFHGEMVLLENYSALNYTGLVKILKKYDKRTGDLMRLPFIQKVL-QQPFYTTDLLFK
AtSPX2  YIIRLK----------ELKDQVAKAK-----------NSNEEMINIKKEIVDFHGEMVLLMNYSALNYTGLAKILKKYDKRTGALIRLPFIQKVL-QEPFFTTDLLNT
StSPX5  YIIHLK----------VLKERVAEMR-----------KSNVEVNRLGRDIYSLENYSALNYTGVVKILKKYDKLSGELLRLPFIQKVL-AEPFFETEVLNK
StSPX1  FIIRFQ----------ELKERIERVKEKSDEDGVLATNNEFSEEMMNIRKDFVSIHGEMVLLKNYSLNFAGLVKILKKYDKRTGGLLRLPFTQVAL-HQPIFTTEPLTR
StSPX4  FIIRLQ----------ELKERIERVKEKSGKDGVLTKDSEFSDEVMGIRKDFVAIHGEMVLLKNYSSLNFAGLVKILKKYDKRTGRLLRLPFTQLAL-DQPFFTTEPLNR
AtSPX4  FVIRLQ----------ELKERIEQVKEKNGE----FASESEFSEEMMDIRDLVTIHGEMVLLKNYSSLNFAGLVKILKKYDKRTGGLLRLPFTQLVL-HQPFFTTEPLTR
OsSPX4  YVIRLQ----------VLKERIERVKAKKNG--AFTSKSEFTEEMLEIRKAFVIIHGEMILLQTYSSSLNFAGLVKILKKYDKRTGGLLSLPFTQRAR-HQPFFTTEPLTR
C-C     ::*: :               ::                               : : .* :*****:** .** :*: *: :*****.**  :   *       .. :: *:: :
```

```
         230       240       250       260       270       280       290       300       310       320       330
    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OsSPX5  LVRDCEAIMDAVFFSLPSPSAAA--------------------------AAAARAAAEQAIFRNTV-AALLTMQEVRSGSSTYGHFSLPPMTP---LPDS-----DWLIQ
OsSPX6  MVRECEAMMAEAA---VLPAEAA-------------------------PEALAAAAEHGIFRNTV-AALLTMEDVRRGGSTHGRHSLPPLT---LPDS-----DWLR-
OsSPX3  LVRDCEATMEAIFTSSVATTAMAGDRRTWKGCSG-------------DAGMAPMADQQGIFRNTV-AALATMKELRSGSSTYGRFSLPPMAAPASPES-----DVLQ-
StSPX2  LVKECESTIDSVFFP----PCDVER-----------------------GLVLGSGVAGEGIFRITV-AALLTMQEIRKGSSTYGRFSLPPLN---LPDQ-----SDVIQ
AtSPX3  LVREWETTMDAVDPVKVAEAEGYER-----------------------CAAVTSAAAGEGIFRNTV-AALLTMKEMRRGGSTYSAFSLPPLN---ISDS-----DNVLR
OsSPX1  LVKQCEAMLDQLLPSNELSVSSEDGRGDSTNEDKPSNPSSS-LVNGGTIPELDEIEYMESMYMKGTV-AALRSLKEIRSGSSTVSAFSLPPLQGDSSPEE-----QQELW
OsSPX2  LVKECEEMLDQLMPTNEHSVASEGKDDSEGEEKGSKPSSSSANGGAVPGEAEAEDERSTDMKSTVTAALRALREIRSGSSTVSVFSLPPLHGSNGQDE-----PGR--
StSPX3  LVKECENLIDRLFPIINTSCITAEADGNETSTSGNNKNDG----LLRSPKELTEIEYIESLYMKSTI-SALRALKEIRSKSSTVSAFSLPPLQISGPED------TWN--
AtSPX1  LVKESEAMLDQIFPANETESEIIQA-----------------------ELSEHKFMESLHMKSTI-AALRVLKEIRSGSSTVSVFSLPPQLNGLDE------TWK--
AtSPX2  FVKECEAMLDRLFFSNKSRNLDEEGEPTTSGMVKTGTDDSE---LLRVPKELSEIEYMESLYMKSTV-SALKVLKEIRSGSSTVSVFSLPPLPASGLEDD-----SWKK-
StSPX5  LVKECDTLLSLLYQTEPLKVAGAGGGGSGGERP-----------VKVPQDLAEIKNMENMYLRLTY-SALRVLQEMRSRSSTVSMFSLPPMNTNALDN------VWKN-
StSPX1  LVHECEANLEVLFFPLEAEVVESAAPADEQMGATSSYASN--------VLPDTTFPQGEENGDIYRSTL-AAIRAIQGLKKASSTHNPLSFSSISGNQDNDSP-GAVTAED-
StSPX4  LVCECEEENLELLFPLEAEVVESDATAEEPKRGTTSYATN--------VSP--ALPLGEENVDIYRSTL-AAIKAIQGLKKASSTYNPLSFSYIFGNQDNDST-GAITAEN-
AtSPX4  LVRECEANLELLFPSEAEVVESSSAVQAHSSSHQHNSPR--------ISAETSSTLGNENLDIYKSTL-AAMRAIRGLQKASSTYNPLSFSSLLQNEDDE----TVTAEN-
OsSPX4  LVRECEANLELLFPIEAEVVELESASSS-AKLQPQNDDAAS-------HDP--ASSVDVETSDVYRSTL-AAMKAIQGLRKASSTYNPLSLARFFHGEDGEACSGAITSESD
C-C     :*  : :   :                                      : *  :*: :. ::  *** .  *:. :            :
```

```
         340       350       360
    ....|....|....|....|....|....|.
OsSPX5  --SVQPPPPPPPSSPLIIPT----------------
OsSPX6  --SFQPPSPIPIQ-----------------
OsSPX3  --SIRSDPHLKENGRIPSLQFFYA-----------
StSPX2  SQSLQLNSPIAIP-----------------
AtSPX3  --SLHLSSPIPIP-----------------
OsSPX1  NKIPVIEQAAK-----------------
OsSPX2  -----------------
StSPX3  -KIPILEQAAK-----------------
AtSPX1  -KIPLLEQEAK-----------------
AtSPX2  -KVGVLEQVAK-----------------
StSPX5  --APVVIQEAK-----------------
StSPX1  -SDSDSLVASQNGEEMNHEVSSSPK----------
StSPX4  --SDSDQSVASQSDKETDQEDSHPPE----------
AtSPX4  --SPN-----SGNKDDSEKEDTGPSH----------
OsSPX4  SYSDSQIEDAEDDDKEVQSREQNTAQNAAEGQPRDE
C-C
```

Figure 3.4 Amino acid sequence alignment and phylogenetic tree of potato (St), rice (Os) and *Arabidopsis* (At) SPX-exclusive family proteins.

The 5 identified potato SPX-exclusive family protein sequences were aligned with those already known from rice and *Arabidopsis* using Clustal X (Higgins *et al.*, 1996; Thompson *et al.*, 1997). Conserved amino acid residues from phosphate binding cluster and lysine surface cluster (Wild *et al.*, 2016) are highlighted in yellow and turquoise, respectively (A, C-C: Clustal Consensus). Phylogenetic tree shown overleaf was generated for these homologues using tree view (Page, 1996) and potato homologues are marked in red (B).

71

Expression profile of SPX domain-exclusive family members during Pi starvation response

As seen in Figure 3.3, the plant P content decreased significantly during the first week of Pi starvation treatment. Therefore, for gene expression profile characterization, both shoot and root total RNA was extracted on the 1st and 7th day of Pi starvation treatment, from 3 individual potato plants either grown on sufficient Pi (300 μM) or low Pi (10 μM) nutrient. The quality of extracted leaf and root RNA samples were examined using Agilent RNA 6000 nano assay (Figure 3.5) before further quantitative experiments. The simulated gel results show that all but one of the RNA samples displayed 2 distinct ribosomal bands corresponding to eukaryotic 18S and 28S ribosomal RNA (Figure 3.5). This indicates successful total RNA extractions, except for leaf sample of high Pi treated replicate 3 on day 1 (Figure 3.5 A). The extra bands seen from the leaf RNA samples represent 23S and 16S ribosomal RNA from the chloroplast.

In order to verify the RNA sample quality from Figure 3.5 A lane 3, cDNA was synthesized via reverse transcription using the same RNA sample as a template, and a PCR reaction was carried out to amplify a fragment from the *StSPX4* gene (Figure 3.6). The PCR result from Figure 3.6 shows a clear band corresponding to the expected amplicon of 274 bp, indicating the original RNA sample was intact and the degradation result seen from Agilent assay could be an artifact caused during sample handling. cDNAs were subsequently synthesized from extracted RNA samples and used for Q-PCR.
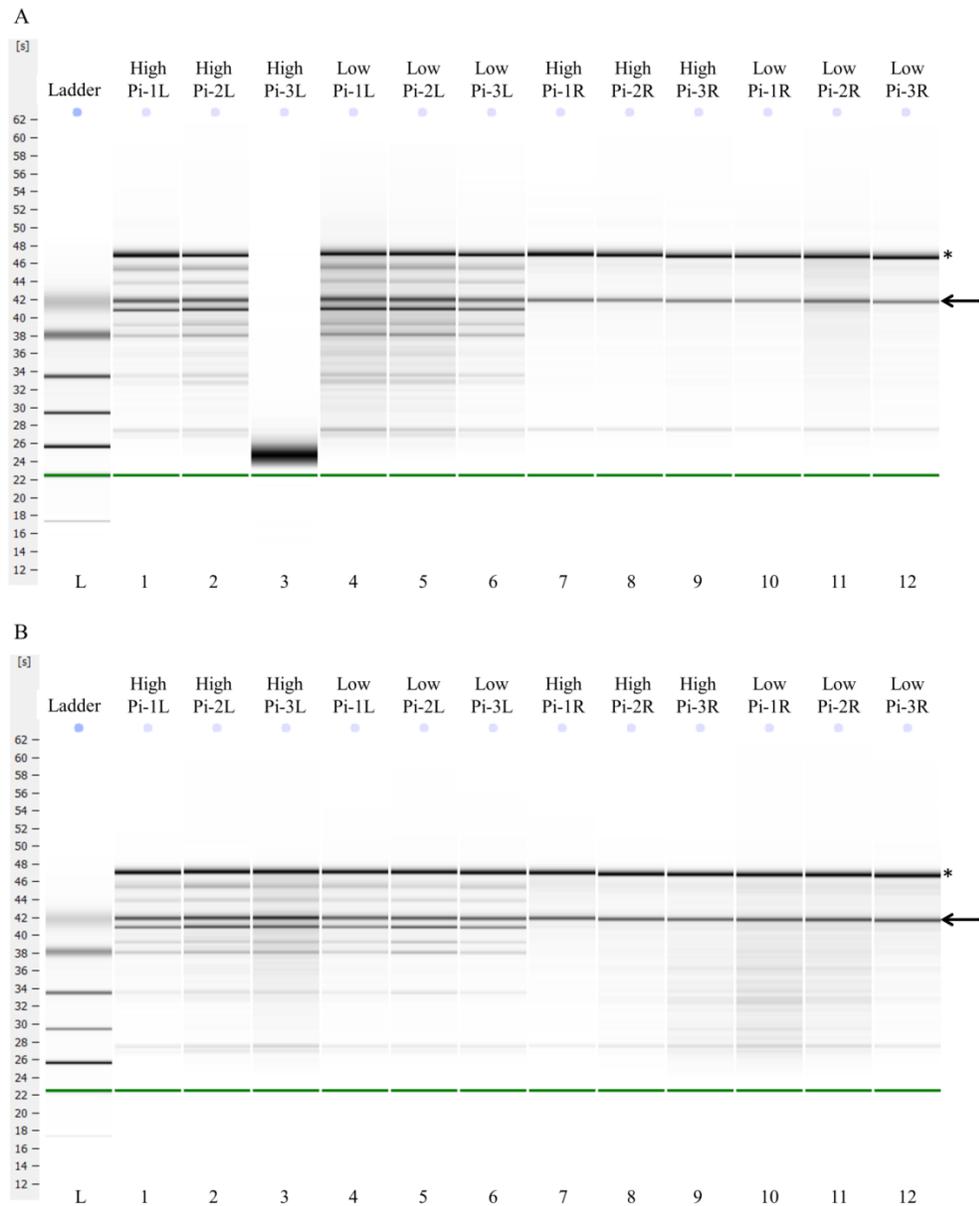
Figure 3.5 Reconstructed gels from plant total RNA quality test using Agilent RNA 6000 nano assay.

Total RNA samples were extracted from day 1 (A) and day 7 (B) after different Pi treatments. The integrity of extracted RNA was tested by Agilent RNA 6000 nano assay using an Agilent 2100 Bioanalyzer. Reconstructed gels were generated by Agilent 2100 Expert software using densitometry plots. (1-3 L: Leaf RNA sample from plant replicate 1-3; 1-3 R: Root RNA sample from plant replicate 1-3). Green line at the bottom of the gels indicates the correctly identified internal lower marker. 28s and 18s rRNAs are indicated by * and an arrow, respectively.
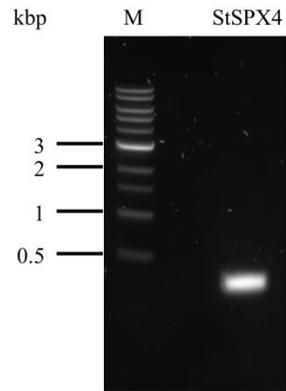
Figure 3.6 RNA integrity test with reverse transcription PCR.

Extracted total leaf RNA sample from day 1 high Pi treated replicate 3 was used as a template for reverse transcription PCR. The expected size of fragment (274 bp) from *StSPX4* gene was amplified using gene specific primers. PCR product was checked on 1.5% agarose gel. (M: NEB 1kb DNA ladder)

Although members from all four plant SPX domain-containing protein families have previously been shown to be involved in Pi regulation pathways (Secco *et al.*, 2012a), the SPX exclusive family members were found to be closely related to Pi starvation responses (Duan *et al.*, 2008; Wang *et al.*, 2009a; Wang *et al.*, 2009b). Previous studies on *Arabidopsis* SPX exclusive family members showed significant transcriptional changes of these genes during Pi starvation response (Duan *et al.*, 2008), whereas the more recent studies in both *Arabidopsis* and rice have demonstrated key roles of members from this family in regulating the PSI genes expression during Pi stress (Puga *et al.*, 2014; Wang *et al.*, 2014). Therefore the gene expression profile on 5 potato SPX exclusive family members (StSPX1-5) during Pi starvation response was studies using quantitative PCR (Q-PCR).

Gene specific primers for each potato StSPX protein corresponding gene were designed using Primer3 (http://primer3.ut.ee). In order to minimize the non-specific amplifications in the subsequent Q-PCR steps, primers were designed either to flank a spliced intron or to contain an intron splicing site, while allowing the production of amplicon size between 100 bp to 150 bp for optimal Q-PCR analysis. Possible primers for each gene were also checked by BLAST searching against potato (*Solanum tuberosum*) genomic database using the EnsemblPlants website (http://plants.ensembl.org/index.html) to make sure no other sequence matches from genomic DNA to each individual primer. The sequences of the primers used are given in materials and methods section 2.2.3, Table 2.2.

The gene encoding potato ribosomal protein L2 (StL2) was selected as a reference gene for Q-PCR for its high expression stability under varying abiotic stresses (Nicot *et al.*, 2005). Primer specificity for each StSPX gene as well as the reference gene StL2 was checked by Q-PCR melt curves (Figure 3.7). The same amount of cDNA synthesized from all leaf and root RNA samples was combined to form a pooled cDNA template for primer specificity tests.
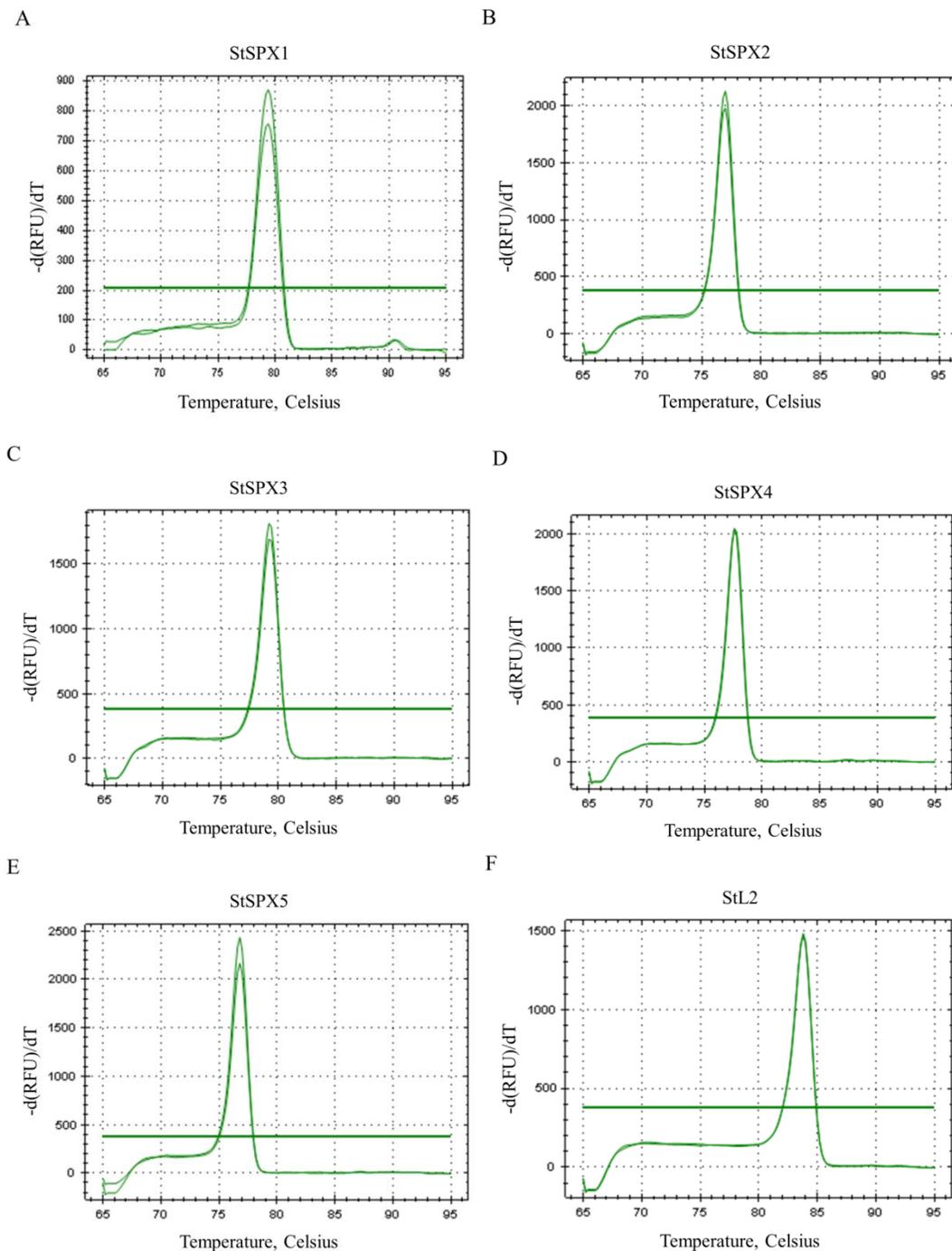
Figure 3.7 Melt curves of Q-PCR products on potato SPX genes and reference gene *StL2*.

Q-PCR was performed using pooled cDNA as a template and designed primers for each potato SPX gene and reference gene at a concentration of 300 nM. Melt curves were produced after the amplification cycles were completed. The change in fluorescence was plotted as a function of temperature to obtain the melt curves for *StSPX1-StSPX5* (A-E) and *StL2* (F). Two technical replicates were included for each gene's Q-PCR reaction.

Melt curves from Figure 3.7 show that when using 300 nM gene specific primers, all the potato SPX genes and the reference gene generated single peaked melt curves, corresponding to a single Q-PCR amplification product for each gene. Two technical replicates from each gene also showed very similar traces of melting curves, indicating the specificity of the chosen primers. The identity of the Q-PCR product from each gene was also confirmed by sequencing before gene expression level changes were measured by quantitative real-time PCR.

Using gene specific primers, the expression level of each *StSPX* gene and the reference *StL2* gene was quantitatively measured using equivalent amounts of template cDNA from each individual sample (Table 3.2). The Q-PCR amplification efficiency for each gene was also examined under the same amplification conditions, using a 3-fold dilution series of all pooled cDNA as templates (Figure 3.8).

Table 3.2 Ct values and Q-PCR amplification efficiency for *StSPX* and reference (*StL2*) genes.

Q-PCR experiments were performed using the CFX96 qPCR System under the control of CFX Manager software. Baseline thresholds for highest amplification efficiencies were reported by the software after amplification cycles finished. Ct (cycle threshold) values show the mean values of two technical replicates and values higher than 40 are not reported (NA). (1, 2, 3: 3 biological replicates)

		StL2	*StSPX1*	*StSPX2*	*StSPX3*	*StSPX4*	*StSPX5*
Threshold Fluorescence		3728.58	3940.45	2619.73	2781.23	1953.46	2866.03
Amplification Efficiency		87.4%	77.3%	66.1%	86.2%	87.3%	75.5%
Ct (cycle threshold) Values							
Day 1	1	28.90	33.67	37.37	35.32	35.99	35.26
Leaf	2	30.85	34.72	NA	37.58	35.37	37.24
300 µM Pi	3	28.29	33.70	36.88	32.93	32.67	32.57
Day 1	1	29.23	34.12	36.04	33.04	36.91	35.05
Leaf	2	30.25	33.53	37.34	35.83	37.25	35.37
10 µM Pi	3	28.88	32.59	39.30	31.84	34.72	32.38
Day 1	1	25.64	31.96	NA	32.94	32.23	34.24
Root	2	25.04	31.53	37.75	32.61	31.13	34.04
300 µM Pi	3	24.67	31.12	NA	30.97	29.40	31.14
Day 1	1	25.91	31.63	35.15	29.71	31.99	30.84
Root	2	25.69	31.01	39.40	29.51	29.89	30.12
10 µM Pi	3	25.23	32.15	34.23	30.30	33.19	30.31
Day 7	1	29.09	33.24	36.57	34.22	34.80	34.95
Leaf	2	28.40	32.50	35.22	33.85	33.31	33.91
300 µM Pi	3	26.17	30.67	34.26	30.73	30.60	28.96
Day 7	1	28.31	32.22	36.80	31.23	32.05	30.91
Leaf	2	28.28	32.17	31.65	28.55	31.06	27.46
10 µM Pi	3	27.89	32.20	36.46	30.99	32.77	30.21
Day 7	1	28.30	30.46	NA	34.38	33.21	NA
Root	2	26.40	32.20	NA	33.13	32.30	34.40
300 µM Pi	3	26.10	31.84	35.68	34.41	32.02	33.99
Day 7	1	27.72	32.84	29.47	28.25	33.84	28.45
Root	2	26.54	31.39	26.36	26.81	NA	27.16
10 µM Pi	3	27.27	31.51	27.98	27.42	35.02	26.84
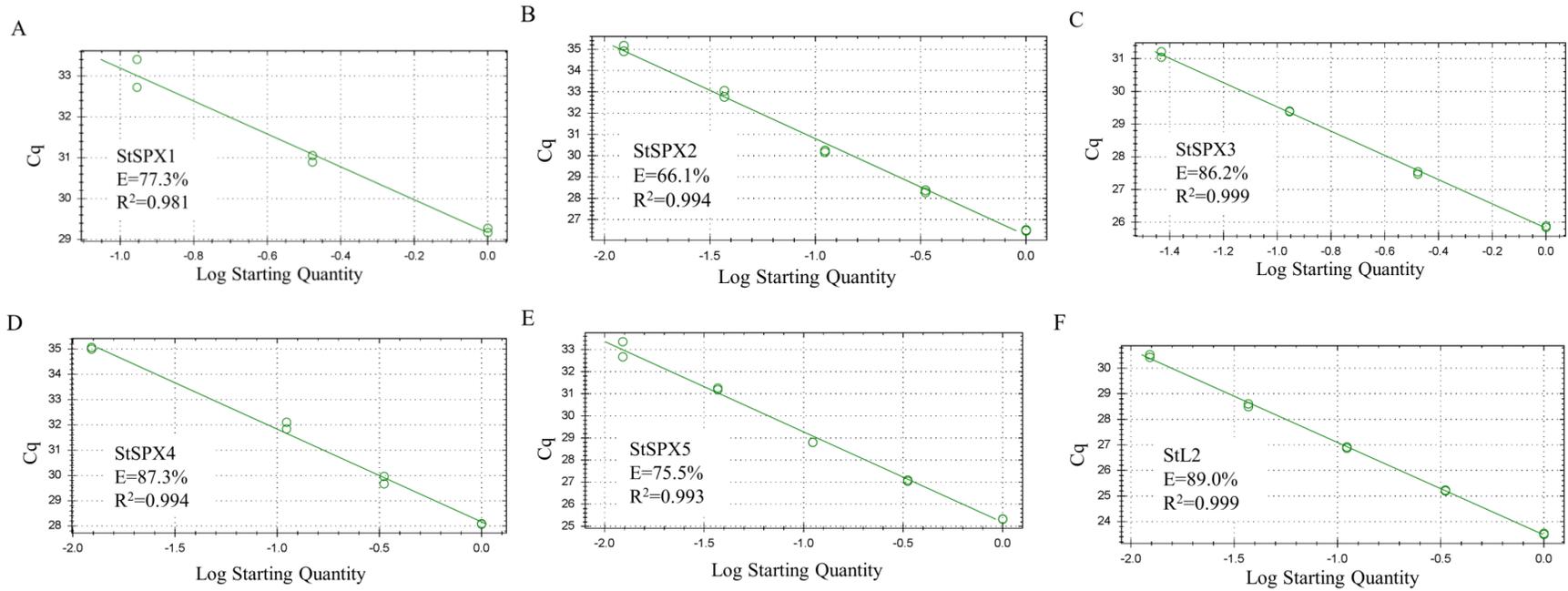
Figure 3.8 Standard curves of quantitative real-time PCR with *StSPX* genes and reference gene *StL2*.

Using a 3 fold dilution series of all pooled cDNA as templates, the Q-PCR amplification efficiency for each gene was examined under the same amplification conditions as for each individual sample. Cq (quantification cycle) values higher than 35 were removed from plotting. Efficiency and fitting coefficient were calculated using CFX Manager software.

Q-PCR results from plant leaf and root samples at different sampling points showed with equivalent amount of RNA input, similar Ct values were obtained from the reference gene *StL2* under different Pi treatments (Table 3.2). It is of notice that small variations of *StL2* expression level were seen at different sampling points. In particular the expression in root on day 1 was higher than that on day 7 (Table 3.2), highlighting the difficulty of finding a stably expressed reference gene under all conditions. However, as the observed *StL2* expression variation is independent of Pi level (Table 3.2), it verifies the suitability of *StL2* as a reference gene for transcriptome analysis under varying Pi availability.

As most of the Ct values from individual samples ranges from 25-35, Ct values higher than 35 were removed from standard curves to reduce the noise generated by low concentrations of initial template. However, although high standard curve fitting coefficients were obtained, lower Q-PCR amplification efficiencies were observed for *StSPX1*, *StSPX2* and *StSPX5* (Figure 3.8). Since altering the dilution series of cDNA or primer annealing temperatures did not improve the amplification efficiency of these *StSPX* genes, the commonly used $2^{-\Delta\Delta C_T}$ method, which requires amplification efficiencies close to 100%, was replaced by its derivation as equation 3.1 (Livak and Schmittgen, 2001), to calculate the relative target gene expression levels.

$$\frac{X_0}{R_0} = \frac{K_X}{K_R} \times \frac{(1 + E_R)^{C_{T,R}}}{(1 + E_X)^{C_{T,X}}}$$

[3.1]

Equation 3.1 Calculation formula of relative gene expression levels [derivation of the $2^{-\Delta\Delta C_T}$ method from (Livak and Schmittgen, 2001)].

X_0: initial number of target molecules; R_0: initial number of reference molecules; K_X: the threshold number of target molecules; K_R: the threshold number of reference molecules; $C_{T,X}$: the threshold cycle for target amplification; $C_{T,R}$: the threshold cycle for reference amplification; E_X: the efficiency of reference amplification; E_R: the efficiency of reference amplification.

Using *StL2* as an internal control (reference gene), the expression levels of *StSPX* genes in leaf and root tissues under different Pi availability were calculated and plotted relative to the expression of *StL2* (Figure 3.9).
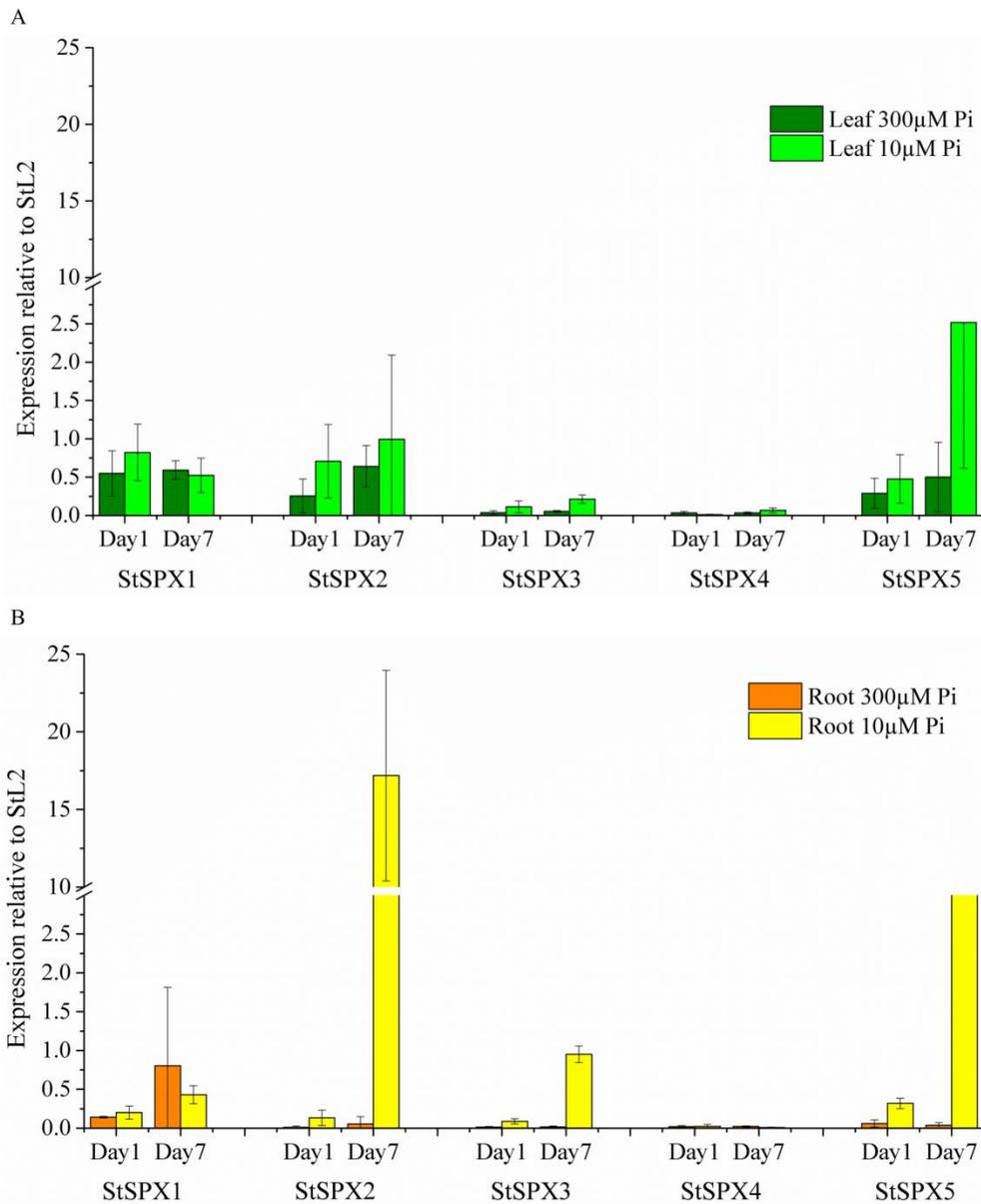
Figure 3.9 Quantitative real-time PCR analysis of relative expression level of *StSPX1-5* in leaf and root samples of hydroponically grown potato.

Potato leaf (A) and root (B) tissue samples were taken on Day1 and Day 7 from high (300 μM) and low (10 μM) Pi treatment. RNA was extracted from each sample and reverse transcribed into cDNA. Q-PCR was performed with *StSPX* gene specific primers and gene expression levels were measured relative to the expression of *StL2*. For each treatment condition, 3 biological replicates each with 2 Q-PCR technical replicates were included. Values are expressed as mean ± SD.

Q-PCR with gene specific primers showed the endogenous *StSPX* genes could be detected in both leaf and root tissues although very low relative expression was observed for *StSPX4* under either Pi availability (Figure 3.9). This suggests that the expression of *StSPX4* is very low and is not Pi starvation inducible in the tissues and under the conditions used in this experiment. For *StSPX1*, the expression was higher in leaf than in root at day 1 and did not change markedly after 7 days of low Pi treatment (Figure 3.9), suggesting the expression of *StSPX1* is not affected by Pi level. In contrast, although expression of *StSPX2* in leaf tissue was not affect by Pi starvation, its expression in root was significantly upregulated after 7 days of Pi starvation ($P<0.05$, n=3; Figure 3.9 B). Expression of *StSPX3* was barely detectable in either leaf or root tissue under sufficient Pi condition, whereas significant upregulation of *StSPX3* in both tissue types expression was seen after 7 days of low Pi treatment ($P<0.01$, n=3; Figure 3.9). Although no significant expression changes were detected for *StSPX5* in leaf tissues, significantly increased expression of *StSPX5* was observed in root on the first day of Pi withdrawal ($P<0.01$, n=3; Figure 3.9 B). This strong upregulation of *StSPX5* in root persisted after 7 days of low Pi treatment ($P<0.05$, n=3; Figure 3.9 B).

It is of note that although tight technical replicates were obtained throughout the Q-PCR experiments, higher variation was seen among the 3 biological replicates under the same treatment conditions (Table 3.2). This also resulted in larger variations for *StSPX* gene expression levels relative to the reference gene *StL2*, and could potentially mask some significant Pi starvation induced expression changes, such as expression of *StSPX5* in leaf tissues after 7 days of Pi limited conditions (Figure 3.9 A).

Although the standard curves generated by template dilution series is traditionally used for Q-PCR amplification efficiency calculations (Rutledge and Côté, 2003), it is sometimes not reliable due to the effect of PCR inhibitors from the most concentrated templates or low copy targets in the most diluted templates. The correction performed for the low efficiencies from standard curves in this Q-PCR study has led to a dramatic change of expression level of *StSPX2* and *StSPX5* relative to the reference ribosomal gene *StL2* in Pi starved potato roots (Figure 3.9 B). However, as higher Ct values were obtained during the amplification of these two genes which corresponds to lower copy numbers, the efficiency values calculated from the standard curves could be affected. Since many other

mathematic algorithms have also been made available for post-run efficiency estimation based on fluorescence signal changes from individual Q-PCR reactions (Graeber *et al.*, 2011), the Q-PCR data was also analyzed using Real-time PCR Miner (Zhao and Fernald, 2005), a post-run calculation algorithm that is found to have better performance on efficiency and Cq values estimation (Tellinghuisen and Spiess, 2014; Ruijter *et al.*, 2013) (Table 3.3). The relative expression levels of five potato SPX genes were calculated using equation 3.1 (Figure 3.10).

Table 3.3 Ct values and Q-PCR amplification efficiency for *StSPX* and reference (*StL2*) genes generated by post-run analysis using Real-time PCR Miner.

Q-PCR fluorescence data from individual reaction wells were analyzed using Real-time PCR Miner (Zhao and Fernald, 2005). Ct (cycle threshold) values show the mean values of two technical replicates and values higher than 40 are not reported (NA). Amplification efficiency for each gene shows the average value from all reactions for the same gene. (1, 2, 3: 3 biological replicates)

		StL2	*StSPX1*	*StSPX2*	*StSPX3*	*StSPX4*	*StSPX5*
Threshold Fluorescence		1616.92	1643.15	1357.78	1404.69	1001.95	1253.98
Amplification Efficiency		87.4%	87.3%	89.3%	88.1%	89.4%	88.4%
Ct Values							
Day 1	1	27.36	32.20	NA	34.09	NA	33.94
Leaf	2	29.51	33.08	NA	NA	34.17	35.61
300 µM Pi	3	26.84	32.29	35.91	31.78	31.49	31.06
Day 1	1	27.80	32.70	35.13	31.88	35.69	34.12
Leaf	2	28.86	32.22	NA	33.51	35.29	NA
10 µM Pi	3	27.51	31.24	NA	30.74	34.10	30.94
Day 1	1	24.21	30.51	NA	31.75	30.95	32.78
Root	2	23.60	30.11	35.23	31.45	29.85	32.60
300 µM Pi	3	23.29	29.74	NA	29.79	28.09	29.58
Day 1	1	24.30	30.20	34.07	28.44	30.70	29.26
Root	2	24.03	29.73	NA	28.37	28.46	28.51
10 µM Pi	3	23.83	30.86	33.10	29.14	32.05	28.83
Day 7	1	27.66	31.84	35.60	33.13	33.76	33.48
Leaf	2	26.92	31.05	34.15	32.68	32.10	32.51
300 µM Pi	3	24.76	29.26	33.20	29.59	29.21	27.33
Day 7	1	26.85	30.82	34.50	30.02	30.84	29.39
Leaf	2	25.62	31.10	30.42	27.98	28.60	25.42
10 µM Pi	3	26.47	30.81	35.42	29.75	31.59	28.71
Day 7	1	26.92	29.10	NA	33.34	32.02	34.65
Root	2	25.00	30.80	NA	32.03	31.05	32.71
300 µM Pi	3	24.71	30.50	34.63	33.31	30.71	32.42
Day 7	1	26.31	31.45	28.28	27.00	32.53	26.85
Root	2	25.17	30.06	25.08	25.50	NA	25.44
10 µM Pi	3	25.89	30.28	26.78	26.19	33.97	25.16
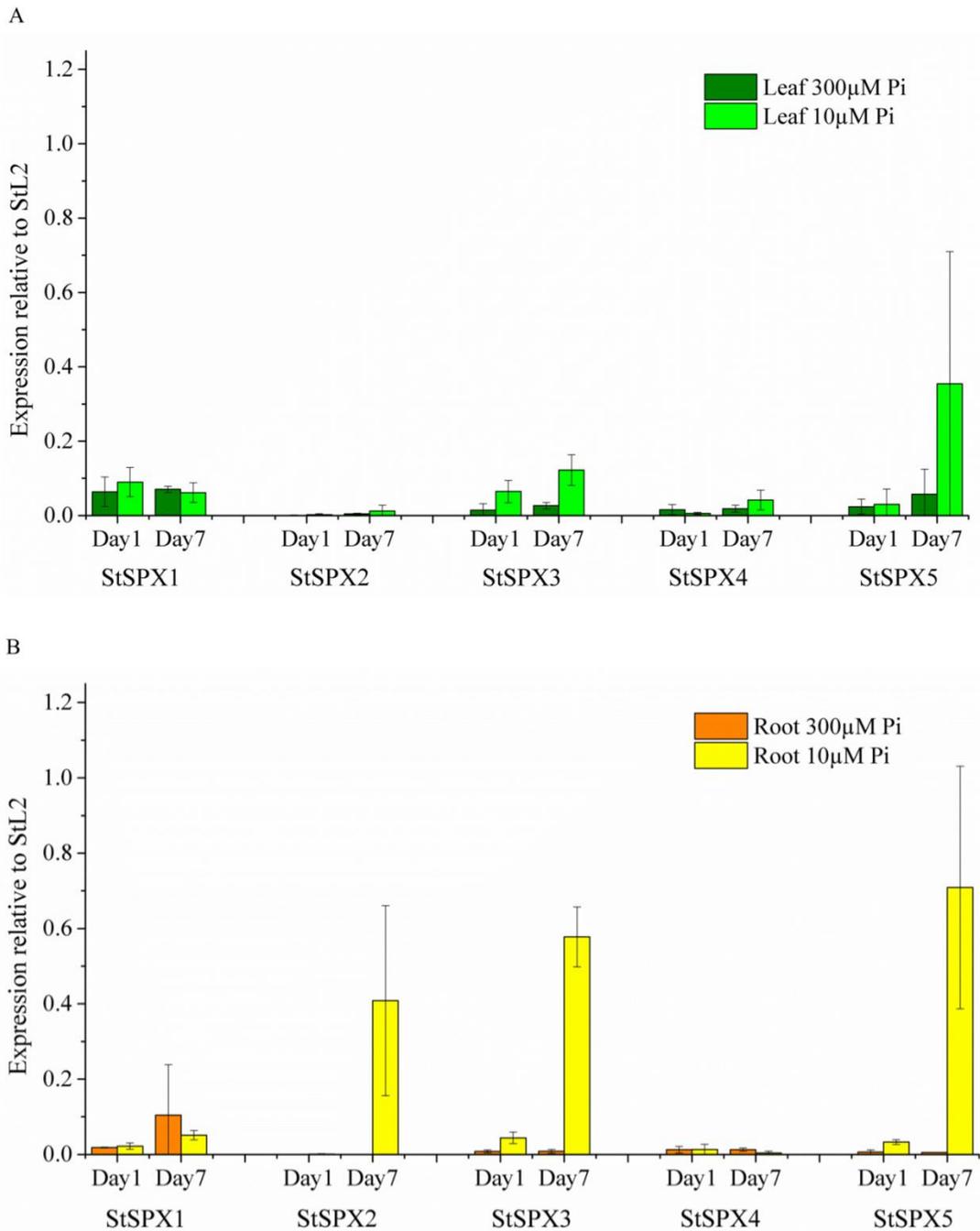
Figure 3.10 Relative expression level of *StSPX1-5* in leaf and root samples using post-run analysis parameters generated by Real-time PCR Miner.

Relative *StSPX* gene expression levels in potato leaf (A) and root (B) tissues were calculated using data from Real-time PCR Miner analysis. For each treatment condition, 3 biological replicates each with 2 Q-PCR technical replicates were included. Values are expressed as mean ± SD.

When the Q-PCR fluorescence signal was analyzed with the post-run algorithm, similar amplification efficiency was obtained for the reference gene *StL2*, whereas the efficiencies for the target *StSPX* genes were much higher than the ones obtained using standard curves (Table 3.3). Although the altered efficiencies from target genes have resulted in much lower relative expression levels of target genes, especially StSPX2 and StSPX5 from Pi starved root (Figure 3.10 B), similar expression profile changes are still maintained using this post-run analysis method. The relative expression of *StSPX4* is still barely detected while *StSPX1* is expressing at a low level relative to *StL2* and the expression of these two genes are not affected by Pi status (Figure 3.10). The expression upregulation of *StSPX2*, *StSPX3* and *StSPX5* detected from the above quantitative analysis are also maintained (Figure 3.10).

In the model plant *Arabidopsis* and a range of other plant species, the expression of a subset of Pi starvation induced genes, including members of SPX exclusive proteins and Pi transporters from PHT1 family, is found to be controlled by PHR1 dependent signalling pathways (Duan *et al.*, 2008; Karthikeyan *et al.*, 2009; Wu *et al.*, 2013). The central transcription factor PHR1 has been demonstrated to regulate target gene expression by specifically recognizing and binding to an imperfect palindromic sequence GNATATNC which is designated P1BS (PHR1 Binding Sequence) (Rubio *et al.*, 2001). In order to investigate whether the same regulation mechanism is also at work in potato Pi starvation responses, the presence of this P1BS cis-element was examined up to 1 kbp upstream of *StSPX* genes. Identified P1BS motifs were mapped according to their upstream positions (Figure 3.11).
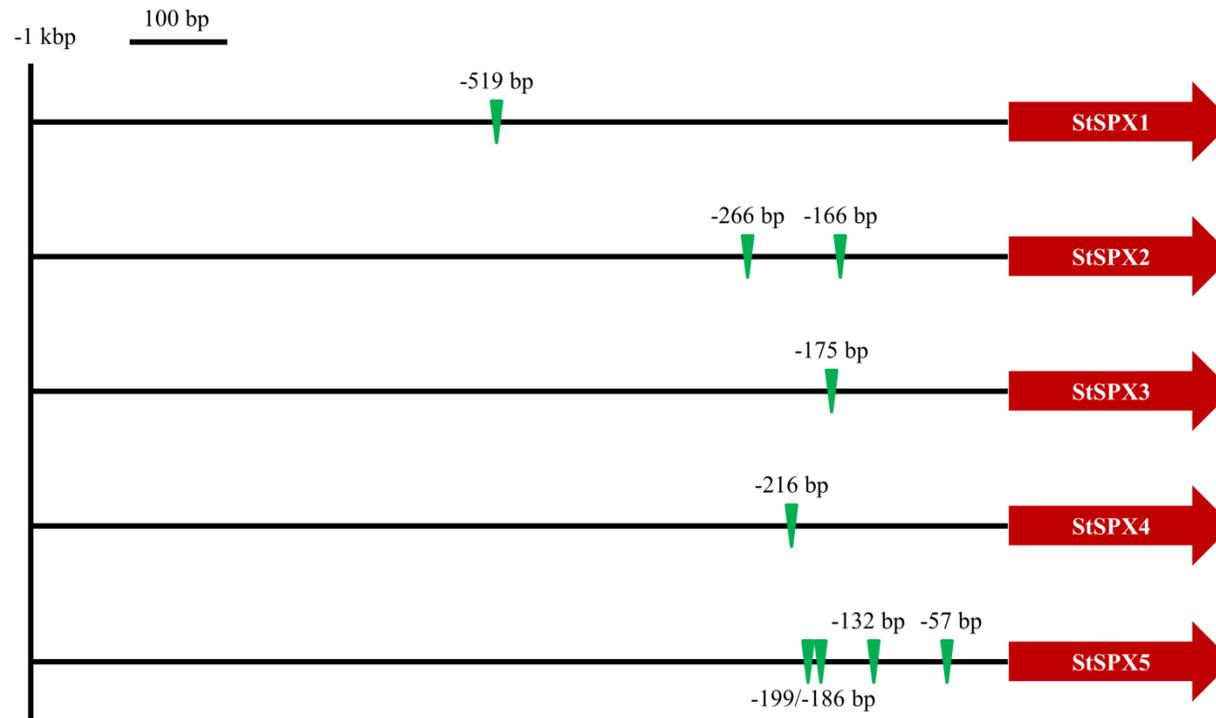
Figure 3.11 Prediction of P1BS motifs upstream of potato SPX genes.

The position of P1BS motifs (GNATATNC) were mapped up to 1 kbp upstream of potato SPX exclusive family genes. The start of gene coding regions are indicated by red arrows and the P1BS motifs are represented by green wedges with corresponding upstream positions labeled in base pairs (bp).

Mapping of P1BS cis-element to 1 kbp upstream of potato SPX exclusive family genes showed this motif is found in the promoter of all *StSPX* genes, mostly within 500 bp distances from the coding regions (Figure 3.11). Multiple P1BS sites were also identified in the upstream sequences of *StSPX2* and *StSPX5* (Figure 3.11). As the expression level of *StSPX2* and *StSPX5* was also shown to be highly upregulated during Pi starvation response (Figure 3.9 B, Figure 3.10 B), this suggests the possibility of *StSPX* genes being regulated by central transcription factors, possibly potato homologues of AtPHR1. Protein amino acid sequence search against *Solanum tuberosum* (tax id: 4113) database using BLAST identified 5 predicted AtPHR1 homologues (Figure 3.12).

Alignment results of these five predicted potato homologs of AtPHR1 showed high conserved consensus with AtPHR1 in the Myb_DNA-binding domain and predicted coiled-coil domain (Figure 3.12), suggesting they are likely to recognize the P1BS cis-element and possibly bind as dimers (see Chapter 5 for dimerization details).

Figure 3.12 Amino acid sequence alignment of *Arabidopsis* PHR1 and predicted potato homologs.

The 5 predicted potato homologs protein sequences were aligned with *Arabidopsis* PHR1 using Clustal X (Higgins *et al.*, 1996; Thompson *et al.*, 1997). Myb_DNA-binding domain and predicted coiled-coil domain from AtPHR1 (Rubio *et al.*, 2001) are highlighted in turquoise and yellow, respectively.

3.3 Discussion

In the laboratory nutrient controlled conditions, potato plants showed obvious morphological changes such as the frequently observed Pi starvation induced shoot growth decrease (Hammond *et al.*, 2003; Wasaki *et al.*, 2003; Calderon-Vazquez *et al.*, 2008; Hammond *et al.*, 2011) as well as physiological changes including reduced phosphorus content in Pi starved leaf tissues (Sato and Miura, 2011; Shin *et al.*, 2004; Hammond *et al.*, 2003). These clear indicators of Pi reduction also validated the Pi starvation condition for the later on gene expression studies. Previous Pi starvation studies on potato plants using a nutrient film technique (NFT) hydroponic system found that although Pi nutrient withdrawal has a delayed effect on plant shoot dry weight, it results in a rapid and significant total leaf P content drop on the first day (Hammond *et al.*, 2011). However, in our study, instead of completely removing Pi nutrient from the growth solution, the hydroponically grown potatoes were provided with Pi nutrient at high (300 µM) and low (10 µM) concentrations that match the K_m ranges of low and high-affinity Pi transporters, respectively (Mitsukawa *et al.*, 1997; Rae *et al.*, 2003; Raghothama and Karthikeyan, 2005). Interestingly, we found that when a lower amount (10 µM) of Pi nutrient is available, in contrary to the previously found total P decreasing on first day, the hydroponic grown potatoes managed to maintain unchanged P content in leaf on the first day of Pi starvation, whereas both total and inorganic P content in leaf tissues decreased significantly within the first week of Pi starvation. Moreover, the same response of P content changes was also observed in potato root tissues, with the decrease of inorganic phosphate contributing to most of the tissue phosphorus content changes. This suggests that when subjected to a low environmental Pi condition, the direct uptake of Pi nutrient through the high affinity Pi transporters such as StPT1, StPT2 (Leggewie *et al.*, 1997; Gordon-Weeks *et al.*, 2003) is capable of maintaining the whole plant P content for at least one day before intracellular Pi concentration decreases.

At the same time, although variation of expression were observed among biological replicates under the same treatment conditions, possibly caused by differences among individual plants, Pi starvation induced gene expression changes were still clearly seen among these potato SPX-exclusive family members. Using Potato Oligo Chip Initiative (POCI) arrays, previous transcriptome studies on potato leaf tissues have demonstrated

significant Pi starvation induced expression suppression of a gene picked up by a *Medicago truncatula* SPX-MFS probe (Hammond *et al.*, 2011). However, Q-PCR data from this chapter suggests that under Pi deprivation conditions, the gene expression changes of the SPX exclusive genes in leaf tissues is not significant except for a small scale upregulation of *StSPX3* after 7 days Pi starvation. On the other hand, a more predominantly low Pi induced transcriptional upregulation of *StSPX2*, *StSPX3* and *StSPX5* is seen in root tissues, suggesting these genes could play key roles in potato root Pi starvation response pathways.

Plant Pi regulation is a complex network which contains both local and systemic signalling pathways as well as cross-talk between metabolites of other nutrients, such as sugar (Yuan and Liu, 2008). Therefore, transcriptome studies using real time Q-PCR requires robust control genes whose expression level is not significantly affected by the tested conditions (Nicot *et al.*, 2005). Previous studies have demonstrated that even the well accepted reference genes such as actin or glycolytic tetramer glyceraldehyde-3-phosphate dehydrogenase (GAPDH) could be rendered unsuitable under certain experimental conditions (Stürzenbaum and Kille, 2001; Radonic *et al.*, 2004). In plants, although some reference genes such as 18S rRNA, ribosomal proteins and actins have been found to maintain a more stable expression level throughout different plant growth stages or under varying treatments (Kim *et al.*, 2003; Volkov *et al.*, 2003), it is also found that suitable reference genes could vary remarkably in different species (Volkov *et al.*, 2003). In potato, the expression stability of seven commonly used reference genes was checked under a series of different conditions, and ribosomal protein L2 and elongation factor 1-α showed more stable expression during different abiotic stresses, especially cold and salt stresses (Nicot *et al.*, 2005), indicating the suitability of these two genes as internal control during the gene expression quantification under Pi starvation conditions. Using the same gene specific primers for *StL2* and *Stef1-α* from Nicot *et al.*, 2005, a higher amplification efficiency was obtained for *StL2* (89%) than for *Stef1-α* (79%) with our Q-PCR samples. Therefore the reference gene *StL2* was chosen as the internal control for *StSPX* gene expression analysis. The Ct values obtained (Tables 3.2 and 3.3) suggest *StL2* is more highly expressed in root (lower Ct values) than shoot, especially at day 1 but that the Pi nutrition status does not have an effect. This is therefore most likely a developmental response. The higher level of expression of *StL2* in day 1 root samples

could lead to an underestimation of SPX gene expression in these tissues but does not change the overall conclusion of the experiment.

Using the standard curve from diluted series of templates, the Q-PCR amplification efficiencies for target *StSPX* genes are relatively low (66%-87%) in spite of attempts to optimize by using different Q-PCR reagents and alternative annealing temperatures being tested. Since acceptable amplification efficiency was obtained from reference gene *StL2* using the same method, this indicates potential low amplification efficiency with certain *StSPX* gene specific primers (Table 3.2). However, since the SPX domain is highly conserved (Secco *et al.*, 2012a), their corresponding genes also share high homology. This gives very low flexibility on the design of primers as gene specific primers are required to discriminate among closely related SPX exclusive family members. As varying the template concentrations in standard curve also induced the inhibition effect from the concentrated templates, the target gene efficiencies could not be further improved using the standard curves method. On the other hand, when the same data was analyzed with Real-time PCR Miner, higher efficiency values (87%-89%) for *StSPX* genes were seen, resulting in much lower relative expression levels of *StSPX* genes. However, as the Ct values higher than 32 are considered much less accurate in the post-run estimation (Zhao and Fernald, 2005), the accuracy of relative expression level calculated from this method might also be affected due to the low expression of most *StSPX* genes.

As the *StSPX* genes are expressed at extremely low level under Pi sufficient conditions (Figure 3.9 and Figure 3.10) but only upregulated during Pi starvation treatment, it is not likely their expression level would become 20-fold higher than the gene encoding the ribosomal protein L2. Therefore, although both Q-PCR data process methods have their limitations regarding the expression of *StSPX* genes, the result obtained from post-run analysis using PCR Miner might be more physiologically plausible. In spite the massively lower relative expression levels of *StSPX* genes obtained using post-run method rendering it impossible to the interpret the fold changes of *StSPX* gene expression under different Pi availabilities, Pi starvation-induced upregulation of *StSPX2*, *StSPX3* and *StSPX5* in potato root tissues are confirmed by both methods. This demonstrates the clear involvement of these three genes in potato Pi starvation response. It is noticeable that large biological variations are seen from both analysis methods, indicating the existence of relatively

strong variations of SPX gene expression from individual plants that might be caused by uncontrollable factors such as leaves being overshadowed during growth or hydroponic aeration disturbance to the root system. Besides, although plant tissues were sampled at the same time of the day for each data point, variations from individual sampling areas could also cause the changes of gene expression among biological replicates.

Although other post-translational process such as protein subcellular localization or degradation (González *et al.*, 2005; Liu *et al.*, 2012) may also participate in the regulation of StSPX proteins, the gene expression profile under Pi starvation conditions was similar between potato *StSPX* genes to those of *Arabidopsis* and rice orthologues (Figure 3.4, Table 3.4) (Duan *et al.*, 2008; Wang *et al.*, 2009a; Wang *et al.*, 2009b), demonstrating strong involvement of SPX domain in Pi regulation among varying species, and that the SPX exclusive family members in the same phylogenetic clade share similar expression regulatory pattern during Pi starvation responses.

Table 3.4 Potato (St), Arabidopsis (At) and rice (Os) SPX exclusive family members gene expression profile during Pi starvation responses.

Potato SPX Gene	Close phylogenetic members (At/Os)	Response to Pi starvation
StSPX1/4	*OsSPX4*	Not obvious response to low Pi stress
StSPX2	*AtSPX3*	Up regulated by low Pi
StSPX3	*AtSPX1/2*	Up regulated by low Pi
StSPX5	*AtSPX1/2*; *OsSPX1/2*	Up regulated by low Pi

Searching for P1BS motif in the upstream region of *StSPX* coding sequences revealed the presence of this cis-element in all *StSPX* genes promoter regions (Figure 3.11). The transcription factor PHR1 has been shown to play a significant role in regulating Pi starvation response by binding to the P1BS cis-element of downstream target genes, including SPX exclusive family members (Rubio *et al.*, 2001; Duan *et al.*, 2008). The identification of P1BS motifs in potato SPX genes thus suggests expression level of *StSPX* genes could be under the control of similar central transcription factors, possibly potato homologues of AtPHR1 (Figure 3.12). Recent studies on *Arabidopsis* phospholipase DZ2 (PLDZ2) also identified a conserved transcriptional enhancer element *EZ2*, containing two

P1BS motifs and critical adjacent sequences in the gene promoter region (Oropeza-Aburto *et al.*, 2012). Subsequent co-expression analysis of *PLDZ2* revealed that a variant of the P1BS motif is significantly over-represented in the promoter region of genes co-expressed with *AtPLDZ2*, including *AtSPX1* and *AtSPX3* and other genes responsible for Pi recycling (Acevedo-Hernández *et al.*, 2012). This demonstrates the great significance of cis-element P1BS in plant Pi regulation and suggests the possibility of identifying genes involved in crop plant Pi regulation by mapping of P1BS motifs in gene promoter regions. As AtPHR1 has been previously suggested to interact with DNA as a dimer (Rubio *et al.*, 2001) and the *in vitro* DNA binding assays in this project also proved the AtPHR1-DNA interaction can be stabilized by tandem P1BS sites (Chapter 5), it suggests genes with multiple P1BS motifs are more likely to be highly upregulated under Pi starvation conditions. This also corresponds to the marked Pi starvation induced expression upregulation of *StSPX2* and *StSPX5*, whose promoter regions harbor multiple P1BS motifs.

Compared to the quite extensive study of SPX domain-containing proteins and their involvement in Pi homeostasis in the model plant *Arabidopsis*, the investigation of SPX domain-containing proteins in crop plants is scarce, except for some experiments in rice (Wang *et al.*, 2009a; Wang *et al.*, 2009b; Secco *et al.*, 2010; Lin *et al.*, 2010). The identification of 14 SPX domain-containing proteins in potato (*Solanum tuberosum*) and their classification into the 4 defined plant families showed the conservation of SPX domain proteins in this major crop plant. By studying the gene regulation pattern of potato SPX exclusive family members under different environmental Pi availabilities, results from this chapter also show the great involvement of SPX proteins in Pi regulation in this agriculturally important species. A previous study on potato Pi starvation responses has demonstrated similar total leaf P content decrease as observed in this study as well as significant expression changes of up to 1659 genes in leaf tissues (Hammond *et al.*, 2011). However, no Pi starvation response data was available from potato root tissues. As our results have demonstrated the P content in root also dropped down significantly after Pi starvation treatment and most of the significant *StSPX* gene expression changes were observed in root, it is important to incorporate the root tissues when analyzing potato Pi starvation responses. Apart from the observed gene expression changes of StSPX proteins during Pi starvation responses, the transcriptome analysis from potato leaves has shown significant expression change of a potential potato *SPX-MFS* gene (Hammond *et al.*,

2011). Therefore, further investigation on the regulation patterns of these SPX-MFS proteins as well as other SPX domain-containing proteins in potato root tissues would help the understanding on how potato plants adapt to Pi limited conditions, and also provide valuable information on breeding potato plants with higher Pi use efficiency.

In order to develop a better understanding of the physiological function of this important SPX domain, extensive overexpression and purification trials on SPX domain from *Arabidopsis* proteins were carried out in the next chapter.

Chapter 4   Expression of *Arabidopsis* SPX protein domains

4.1   Introduction

Results from chapter 3 have shown significant gene expression upregulation of some SPX exclusive family members in the agriculturally important crop plant potato during Pi starvation. Although the importance of the SPX domain for Pi regulation has been demonstrated in different plant species [reviewed in (Secco *et al.*, 2012a)], at the outset of this project little information was available on the structure of this protein domain or the molecular mechanism of their involvement in Pi regulation. In this chapter, multiple expression constructs for SPX protein domains were screened under a range of conditions in the attempt to find a suitable construct which would allow biochemical and structural studies on this highly conserved domain.

As the Pi starvation induced expression profile changes of potato SPX proteins were only made available at a later stage of this project, proteins from this species were not used for the expression for structural and functional studies. On the other hand, many SPX domain-containing proteins in the model plant *Arabidopsis* had been shown to be involved in Pi regulation prior to this project (Duan *et al.*, 2008; Stefanovic *et al.*, 2007; Kant *et al.*, 2011; Liu *et al.*, 2015) and their cDNA clones encoding the full length proteins are available (Arabidopsis Biological Resource Center). Therefore, various versions of *Arabidopsis* SPX protein domains were constructed for expression in both plant (Tobacco and *Arabidopsis*) and *Escherichia coli* (*E. coli*) with the goal of obtaining correctly folded SPX protein domains in a soluble form in quantities sufficient for subsequent purification and functional studies.

4.2 Results

4.2.1 Expression of AtSPX1 in tobacco leaves

Although studies on *Arabidopsis* SPX proteins have demonstrated all members from the SPX exclusive family show significant expression changes during Pi starvation responses, AtSPX1 has been found to be of the most significance for exhibiting regulatory functions. Apart from the strong Pi starvation induced expression, AtSPX1 is also found to be nuclear localized and to increase the expression of PSI genes *ACP5*, *RNS1* and *PAP2* when over-expressed (Duan *et al.*, 2008). As tobacco (*Nicotiana benthamiana*) leaves have previously been shown to be an effective heterologous expression system to obtain functional SPX domain-containing proteins (Zhao *et al.*, 2009; Wege and Poirier, 2014), the expression of full-length native AtSPX1 was tested in tobacco leaves.

Expression constructs for expressing N- and C-terminal His$_6$ tagged AtSPX1

To transiently express AtSPX1 in tobacco leaves, the binary plasmid pEAQ-*HT* was used to allow the enhanced expression of gene of interest (*AtSPX1*) with a flanking CPMV-*HT* cassette, as well as the fusion of either a C- or N-terminal His$_6$ tag (Sainsbury *et al.*, 2009). The full-length *AtSPX1* coding sequence (TAIR, At5g20150) was sub-cloned into this pEAQ-*HT* binary vector using restriction enzymes to generate both N- and C-terminal His$_6$-fused expression vector pEAQ-*HT*-His$_6$-AtSPX1 and pEAQ-*HT*-AtSPX1-His$_6$ (Figure 4.1). Expression constructs were confirmed by sequencing before being used for subsequent expression trials.
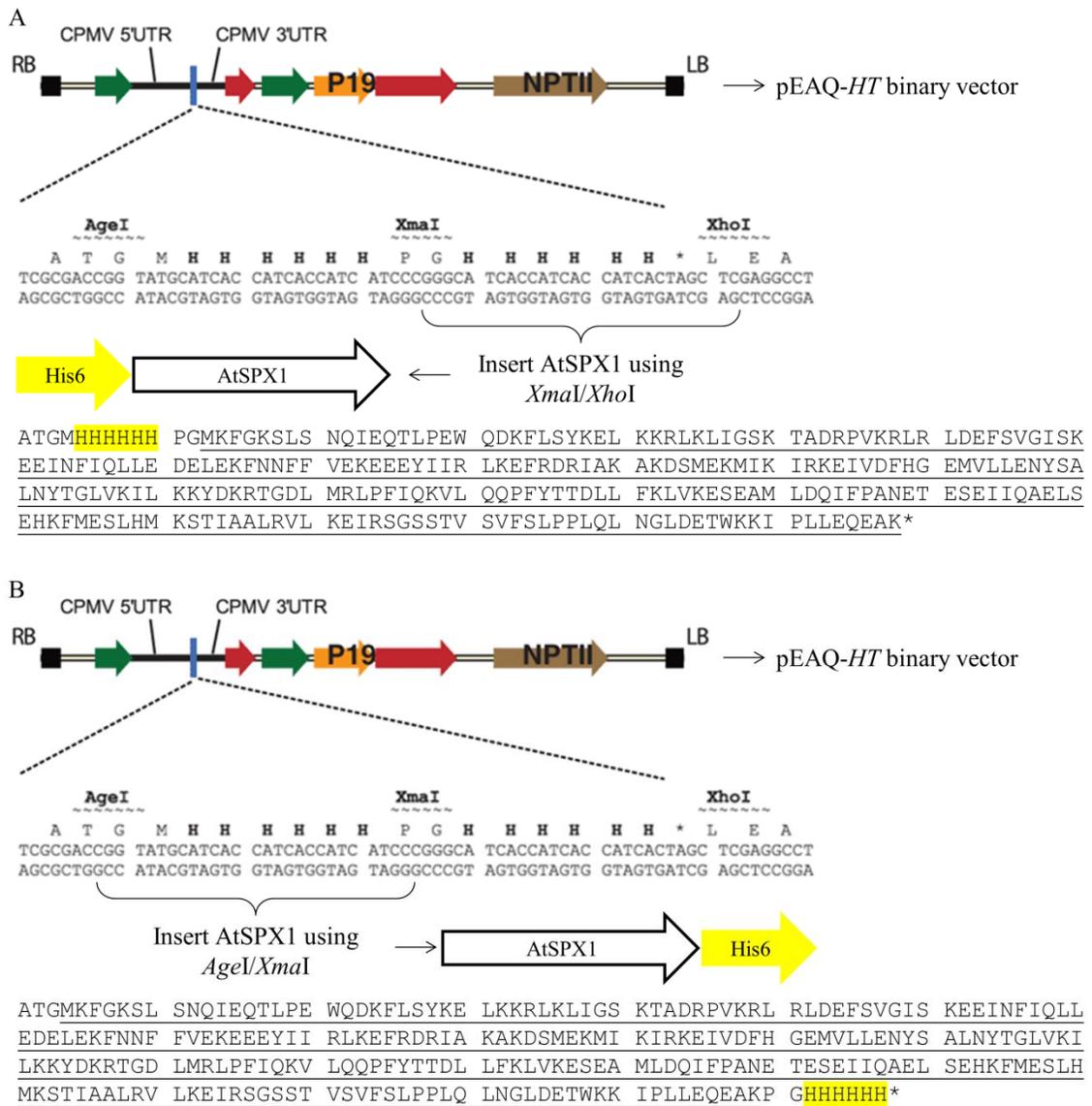
Figure 4.1 Expression constructs and protein sequences for N- and C-terminal His₆-fused AtSPX1.

Gene encoding full-length AtSPX1 was sub-cloned into pEAQ-*HT* binary vectors that contain a CPMV-HT cassette (a modified 5'-untranslated region (UTR) and the 3'-UTR from Cowpea mosaic virus (CPMV) RNA-2) flanking inserted genes for expression (Sainsbury *et al.*, 2009). Available *Xma*I / *Xho*I sites and *Age*I / *Xma*I sites were used to generate N-terminal and C-terminal His₆ fused AtSPX1 constructs pEAQ-*HT*-His₆-AtSPX1 (A) and pEAQ-*HT*-AtSPX1-His₆ (B), respectively. *AtSPX1* sequences are marked with black underline and His₆ tags are marked with yellow highlight. (P19: the suppressor of silencing; NPTII: neomycin phosphotransferase II gene, conferring resistance to kanamycin; LB: T-DNA left border; RB: T-DNA right border).

99

Agrobacteria mediated plant transfection and protein expression detection

For transient expression of AtSPX1 in tobacco leaves, the *Agrobacterium* cells were transformed with the above N- and C-terminal His_6 fused AtSPX1 constructs and subsequently infiltrated into tobacco epidermal cells for expression using the method described in Sparkes *et al.*, 2006. *Agrobacteria* containing the expression vector were tested at a final OD_{600} of 0.1 and 0.5 for infiltration. A pEAQ-HT-GFP construct (Sainsbury *et al.*, 2009) was used as a positive control for *Agrobacterium* transformation and was infiltrated alongside the AtSPX1 expression constructs. Tobacco leaves were collected on the 5th day and 10th day following infiltration and expression of positive control GFP was examined by *in situ* fluorescence imaging as well as western blotting (Figure 4.2).

Examination of the positive control indicated the successful transient expression of GFP protein in tobacco leaves after infiltration (Figure 4.2). Western blotting using the same amount of total extracted protein revealed a slightly higher over-expression level of GFP after 10 days of infiltration compared to 5 days, although the expression increase on day 10 could be underestimated due to the saturated western blotting signal (Figure 4.2 C). In order to examine the expression of N- and C-terminal His_6 fused AtSPX1, total protein was extracted using denaturing extraction buffer containing 1% SDS (Chapter 2 section 2.2.5.1) from infiltrated leaves after 5 and 10 days of infiltration and the presence of target proteins in both soluble and non-soluble extraction fractions were detected by western blotting with anti-His_6 antibody (Figure 4.3, Figure 4.4).
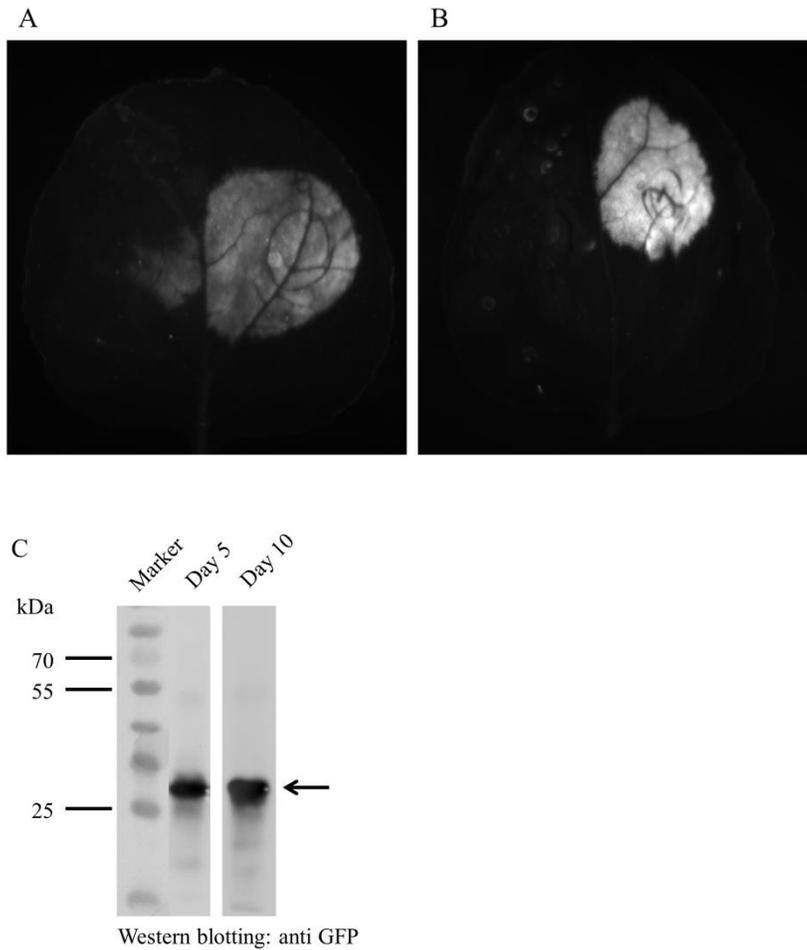
Figure 4.2 Expression of positive control GFP on the 5th and 10th day following tobacco leaf infiltration.

Tobacco leaves infiltrated with *Agrobacteria* containing pEAQ-*HT*-GFP were collected on the 5th day (A) and 10th day (B) of infiltration and *in situ* expression of GFP was visualized using the G-box fluorescence imaging system. Total soluble proteins were extracted from infiltrated leaf samples using 1% SDS and 20 μg proteins were separated on 12% SDS-PAGE gels and transferred to a nitrocellulose membrane for immunoblotting using anti-GFP antibody (C). GFP is indicated by an arrow.
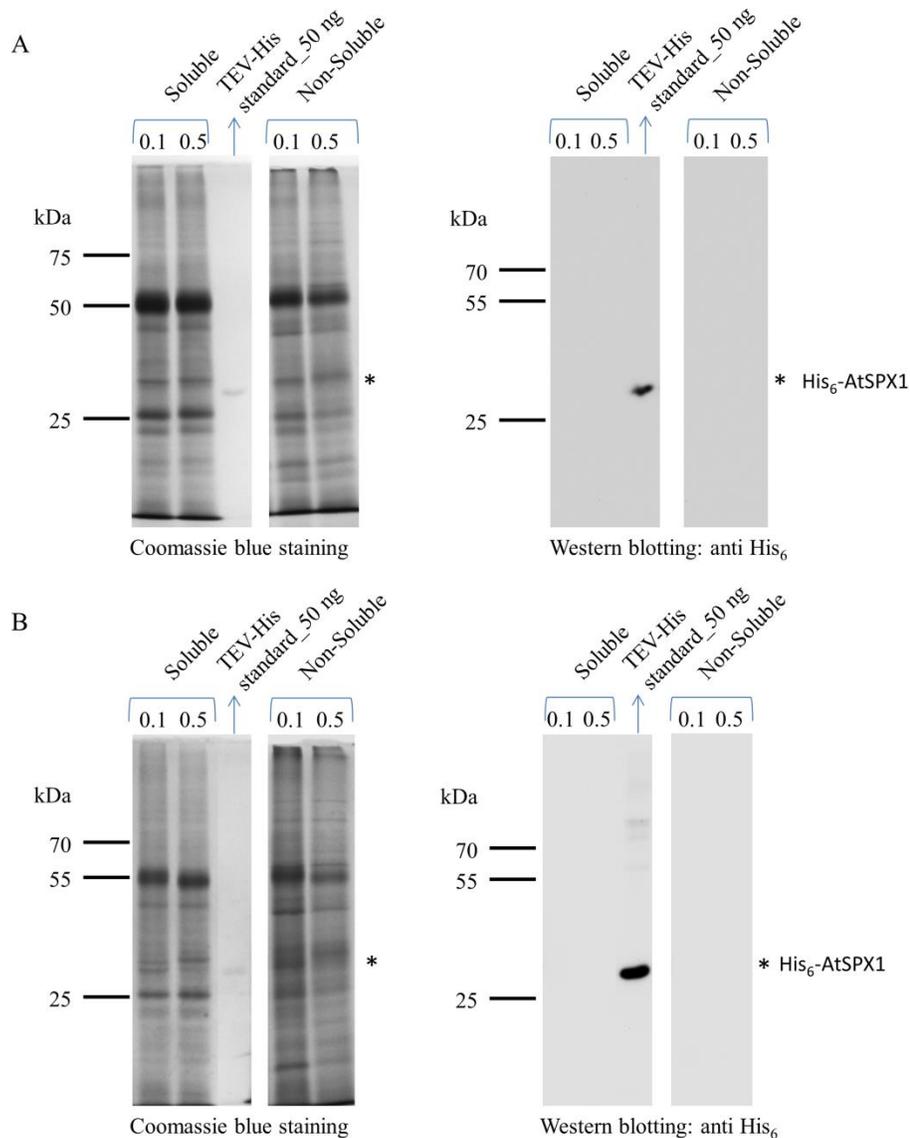
Figure 4.3 Expression of N-terminal fused AtSPX1 on the 5th and 10th day of tobacco leaf infiltration

Tobacco leaves infiltrated with *Agrobacteria* containing pEAQ-*HT*-His$_6$-AtSPX1 were collected on the 5th day (A) and 10th day (B) of infiltration. Total soluble proteins were extracted and 20 µg soluble proteins and equivalent amount of non-soluble fractions re-suspended in SDS-PAGE loading buffer were separated on 12% SDS-PAGE gels. Proteins were stained with Coomassie blue (left panels). An equal amount of proteins were also transferred to a nitrocellulose membrane after SDS-PAGE for immunoblotting using anti-His$_6$ antibody (right panels). The expected size of N-terminal His$_6$ fused AtSPX1 (Theoretical Mw 31.1 kDa) is indicated by *. (0.1: leaf sample infiltrated with *Agrobacteria* at OD$_{600}$ of 0.1; 0.5: leaf sample infiltrated with *Agrobacteria* at OD$_{600}$ of 0.5.)
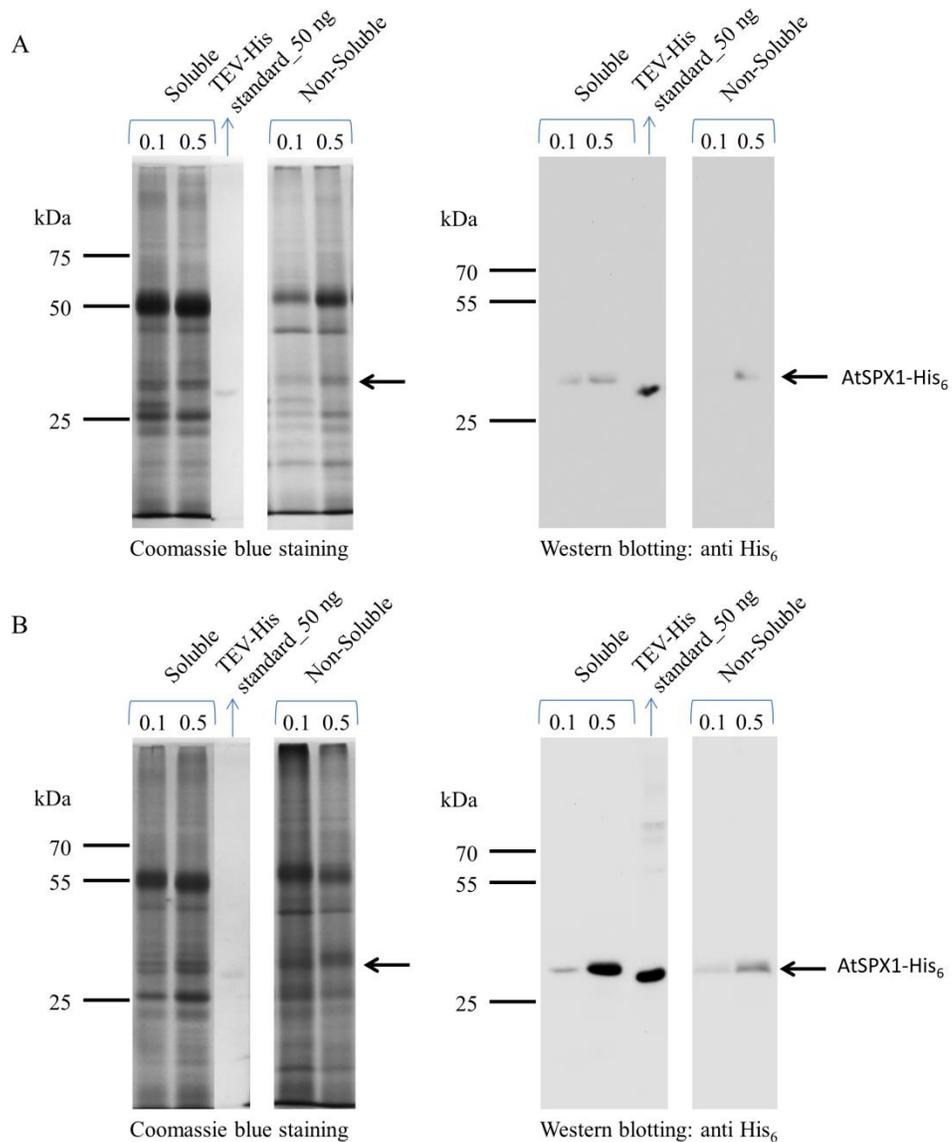
Figure 4.4 Expression of C-terminal fused AtSPX1 on the 5th and 10th day of tobacco leaf infiltration.

Tobacco leaves infiltrated with *Agrobacteria* containing pEAQ-*HT*-AtSPX1-His$_6$ were collected on the 5th day (A) and 10th day (B) of infiltration. Total soluble proteins were extracted and 20 μg soluble proteins and equivalent amount of non-soluble fractions re-suspended in SDS-PAGE loading buffer were separated on 12% SDS-PAGE gels. Proteins were stained with Coomassie blue (left panels). An equal amount of proteins were also transferred to a nitrocellulose membrane after SDS-PAGE for immunoblotting using anti-His$_6$ antibody (right panels). The C-terminal His$_6$ fused AtSPX1 (Theoretical Mw 31.1 kDa) are indicated by an arrow. (0.1: leaf sample infiltrated with *Agrobacteria* at OD$_{600}$ of 0.1; 0.5: leaf sample infiltrated with *Agrobacteria* at OD$_{600}$ of 0.5.)

103

Western blotting results showed successful expression of the C-terminal His_6 fused AtSPX1 in tobacco leaves (Figure 4.4). A higher expression level of target protein was also observed on the 10th day of infiltration using *Agrobacteria* at OD_{600} of 0.5 (Figure 4.4 B). When compared to the standard His_6-tagged protein (TEV-His) western blotting signals, expression under this condition gave a typical yield of approximately 35 µg target protein per gram of fresh leaf tissue.

On the other hand, no expression was observed for the N-terminal His_6 fused AtSPX1 when total extracted proteins were blotted against anti-His_6 antibody (Figure 4.3). Since 1% SDS is shown to be efficient in extracting over-expressed protein from tobacco leaf samples (Figure 4.4), this indicates the N-terminal tag might interfere with the over-expression or correct folding of AtSPX1.

Protein extraction and affinity purification trial

As the C-terminal His_6 fused AtSPX1 was shown to express well, larger scale expression (Chapter 2 section 2.2.1.3) was carried out by infiltrating tobacco leaves with *Agrobacteria* containing pEAQ-*HT*-AtSPX1-His_6 expression construct at OD_{600} of 0.5. Infiltrated samples were collected on the 10th day of infiltration and total proteins were extracted using a native extraction buffer containing 0.1% Triton X-100, plant protease inhibitor and 1 mM PMSF (Chapter 2 section 2.2.5.3). C-terminal His_6 fused AtSPX1 was subsequently enriched via Immobilized Metal Affinity Chromatography (IMAC) using Ni-NTA resin (Figure 4.5).
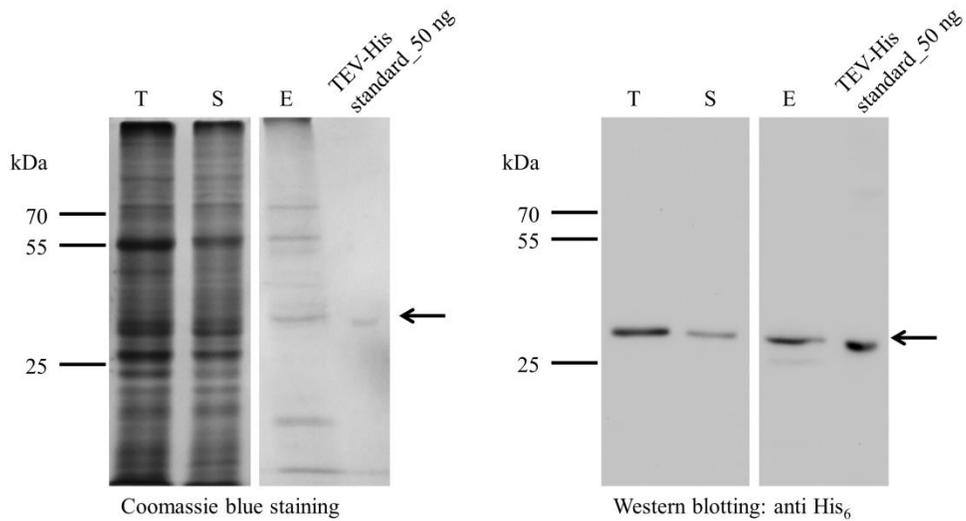
Figure 4.5 SDS-PAGE analysis of immobilized metal affinity chromatography (IMAC) purified C-terminal His$_6$ fused AtSPX1.

IMAC of C-terminal His$_6$ fused AtSPX1 was carried out using Ni-NTA resin on a gravity flow column. An equivalent amount of proteins from total protein extract (T), soluble protein fraction (S) and purification elution fraction (E) was separated on 12% SDS-PAGE gel and stained using Coomassie blue (left panel). Equal amount of proteins were also transferred to a nitrocellulose membrane after SDS-PAGE for immunoblotting using anti-His$_6$ antibody (right panel). The expected size of C-terminal His$_6$ fused AtSPX1 (Theoretical Mw 31.1 kDa) is indicated by an arrow.

Although the amount of C-terminal His_6 fused AtSPX1 that could be extracted with SDS was reasonably high (Figure 4.4 B), the amount of soluble protein decreased when extracted in the native form (Figure 4.5), suggesting a lower extraction efficiency using the native extraction buffer compared to the denaturing buffer containing SDS. Compared to the western blotting signals from standard His_6-tagged proteins (TEV-His), this IMAC purification procedure had a typical yield of 0.3-0.5 μg purified protein per gram of fresh leaf tissue while other contaminations are shown to be more dominant in the IMAC elution fraction (Figure 4.5, left panel). The low amount of recovered target protein after IMAC also indicates possible protein degradation or precipitation during cell lysis and purification.

Given the low yield of target protein and the relatively longer period for protein expression compared to the *E. coli* expression system, the *Agrobacteria* mediated transient expression in tobacco leaves was demonstrated to be not suitable for obtaining sufficient recombinant *Arabidopsis* SPX1 protein for subsequent functional or structural studies.

4.2.2 Expression of AtSPX1 in *Arabidopsis* protoplasts

In order to identify potential *in vivo* interacting partners, AtSPX1 was expressed in *Arabidopsis* protoplasts for *in vivo* pull-down experiments. The coding sequence of full-length AtSPX1 (TAIR, At5g20150) was sub-cloned into a pUC18 based expression vector to generate a C-terminal His_8-GFP fused construct. *Arabidopsis* mesophyll protoplasts were prepared using the 'Tape-*Arabidopsis* Sandwich' method (Wu *et al.*, 2009) and transient expression of AtSPX1-His_8-GFP fusion protein was carried out using varying amount of plasmid DNA (Chapter 2 section 2.2.1.2) (Figure 4.6).

Examination of transfected protoplasts under a fluorescence microscope showed successful transient expression of C-terminal GFP fused AtSPX1. GFP signal from protoplasts was observed in the transfected groups using varying amount of plasmid DNA (Figure 4.6 – 10~50 μg), whereas no signal was obtained from the control group (Figure 4.6 – 0 μg). Based on the percentage of fluorescent protoplasts, the highest transfection efficiency was seen using 20 – 30 μg plasmid DNA, and further increasing the amount of DNA did not improve the transfection efficiency. The observed nuclear localized GFP signal further suggest the transiently expressed AtSPX1 is very likely to have obtained a correctly folded structure and maintained its nuclear subcellular localization as reported before (Duan *et al.*, 2008). However, during the course of this experiment, it was reported that AtSPX1 interacts with *Arabidopsis* transcription factor PHR1 (PHosphate starvation Response 1) in a Pi-dependent manner (Puga *et al.*, 2014). Given the limited amount of protein could be obtained from *Arabidopsis* protoplast transient expression, the purpose of this experiment was primarily to identify *in vivo* interacting partner of AtSPX1 rather than producing large quantity of AtSPX1 protein for subsequent studies. Therefore, this protoplast transient expression experiment was stopped and the research focus was placed on the expression and purification of sufficient amount of recombinant SPX proteins for structural and biochemical studies.
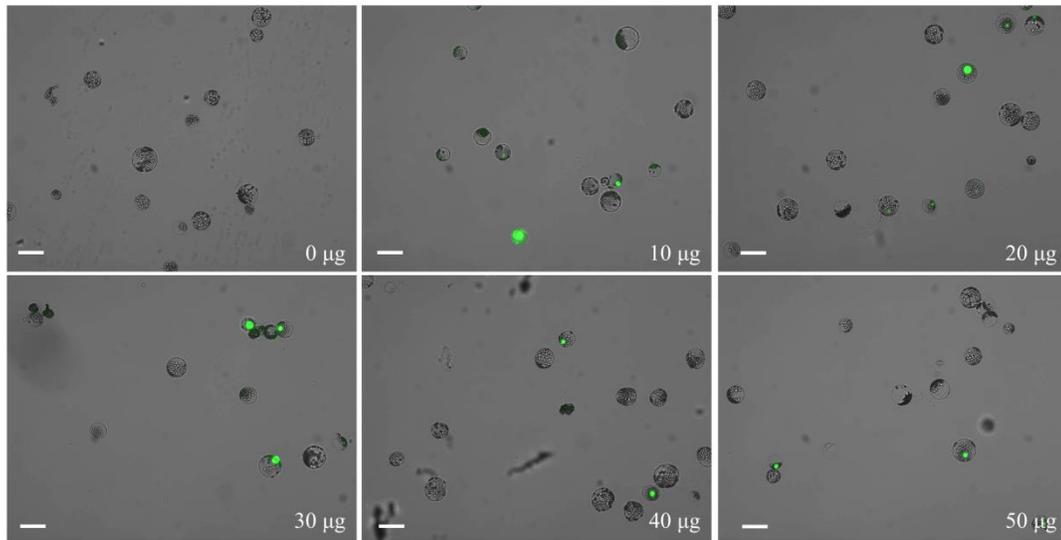
Figure 4.6 Transient expression of C-terminal His$_8$–GFP fused AtSPX1 in *Arabidopsis* protoplasts.

Varying amount of AtSPX1-His$_8$-GFP construct was used for transfection of *Arabidopsis* mesophyll protoplasts prepared using the 'Tape-Arabidopsis Sandwich' method (Wu *et al.*, 2009). Transfected protoplasts were incubated at 25°C for 22 h before being examined under a fluorescence microscope. The scale bar equals 50 μm.

4.2.3   Expression trials in prokaryotic (*E. coli*) cells

Since the plant protein expression system using tobacco leaves was proved to be inefficient for expressing *Arabidopsis* SPX protein, the expression trials were carried out using *E. coli* cells. Expression constructs with different tags and varying versions of *Arabidopsis* SPX domains were tested in order to obtain the soluble SPX protein domain for further investigation.

Choosing and obtaining of SPX domains for expression

As the SPX domain-exclusive family members AtSPX1 and AtSPX3 have been demonstrated to be pivotal positive regulators of plant adaptation to phosphate deprivation (Duan *et al.*, 2008; Secco *et al.*, 2012a), the full-length coding regions of these two proteins were obtained from ABRC (AtSPX1: U21625; AtSPX3: U68223). Analysis of the codon adaptation index (CAI) for expression native coding sequence of AtSPX1 in *E. coli* using JCat (Grote *et al.*, 2005) yielded a value of 0.196, suggesting very different codon employment between *Arabidopsis* and *E. coli* cells. Therefore without changing the amino acid sequences of AtSPX1 or introducing frequently used restriction sites, a codon usage optimized sequence coding AtSPX1 (Appendix 1, done by Prof. Stephen Baldwin) was synthesized and tested for the heterologous over-expression in *E. coli* cells.

In contrast to the well conserved N-terminal SPX domain, the C-terminal regions of AtSPX1 was predicted to be disordered using DisoPred (http://bioinf.cs.ucl.ac.uk/disopred) (Figure 4.7). In order to prevent the protein degradation or precipitation caused by the disordered regions, sequences coding two shorter versions of AtSPX1 containing only the conserved domains (AtSPX1-S1 and AtSPX1-S2) (Figure 4.7) were also included for the expression screen. To create these two truncated versions of AtSPX1, the codon optimized *AtSPX1* coding sequence was used as a template and gene sequences coding amino acid 1-159 (AtSPX1-S1) and 1-181 (AtSPX1-S2) were amplified, cloned and sequenced.

```
  1 MKFGKSLSNQ IEQTLPEWQD KFLSYKELKK RLKLIGSKTA DRPVKRLRLD EFSVGISKEE 60
 61 INFIQLLEDE LEKFNNFFVE KEEEYIIRLK EFRDRIAKAK DSMEKMIKIR KEIVDFHGEM 120
121 VLLENYSALN YTGLVKILKK YDKRTGDLMR LPFIQKVLQQ PFYTTDLLFK LVKESEAMLD 180
181 QIFPANETES EIIQAELSEH KFMESLHMKS TIAALRVLKE IRSGSSTVSV FSLPPLQLNG 240
241 LDETWKKIPL LEQEAK 256
```

Figure 4.7 Amino acid sequence of AtSPX1.

The 256 aa full-length AtSPX1 harbors a conserved N-terminal SPX domain (highlighted in yellow) as well as predicted C-terminal disordered regions (highlighted in green). Two shorter versions excluding the disordered regions, AtSPX1-S1 (1-159) and AtSPX1-S2 (1-181) are marked red and underlined, respectively.

Previous research in *Arabidopsis* has indicated that the PHO1 protein family is also significantly involved in maintaining Pi homeostasis (Poirier *et al.*, 1991; Stefanovic *et al.*, 2007) (Chapter 1, section 1.4.2). Members of this family carry a longer version of the SPX domain at their N-termini due to the longer sequences between the highly conserved sub-domains (Secco *et al.*, 2012a). N-terminal SPX domains from AtPHO1;H1 of different lengths (AtPHO1;H1_1: aa 1-340 and AtPHO1;H1_2: aa 1-384) (Figure 4.8) were selected for expression tests in *E. coli* cells. The C-terminal extensions of AtPHO1;H1 SPX domains were selected following examination of an alignment of the sequences of all AtPHO1 family members (Appendix 6). Each boundary was selected right before a poorly conserved residue to give sequence corresponding to the conserved SPX domain with or without an adjacent C-terminal well conserved region. Plasmids containing the coding region for these two AtPHO1;H1 SPX domains were provided by Prof. Stephen Baldwin.

```
  1 MVKFTKQFEG QLVPEWKDAF VDYSQLKKDL KKIHLFTNGV EKKHTETSLI KTVKSSLGRL  60
 61 SIFGNKGREQ SRVIQVHKKL ASSGSNNDVY ETELLEKIAD DTDAAKEFFA CLDMQLNKVN 120
121 QFYKTKEKEF LERGECLKKQ MDILIELKDA FKQKQANGES TQESKEDDSI SCTISCEYDS 180
181 VRGRTEEMQL QVSCLDNLED NGEEALESLG SEEPIKANNE DSKLTTVSSR VFSCQGKNVK 240
241 IKIPLTNPSR TFSAISYLIN QSSSKKNGPD GGNKLQISKK KLSHAEKMIK GALTELFKGL 300
301 NYLKTYRNLN ILAFMNILKK FDKVTGKQIL PIYLKVVESS YFNISDKVMI LSDEVEEWFI 360
361 KHLAGENRRK AMKYLKPHHR KESHSVTFFI GLFTGCFVAL LAGYIIVAHL TGMYRQHSAN 420
421 TFYMETAYPV LSMFGLLFLH LFLYGCNIFM WRKARINYSF IFELGSKNEL KYRDVFLICT 480
481 ASMSAIAGVM FVHLSLLEKG YSFRQVQVIP GLLLLGFLLI LICPLNIFYK SSRYRLISVI 540
541 RNIVFSPLYK VVMLDFFMAD QLCSQVPMLR NLEYIACYYI TGSYATQDYE YCMRVKYYRD 600
601 LAYAVSFLPY YWRAMQCARR WFDEGETSHL VNLGKYVSAM LAAGTKVAYE KERSLGWLCL 660
661 VVAMSSVATI YQLYWDFVKD WGLLQHNSNN PWLRNQLMLR QKSIYYFSMV LNLVLRLAWL 720
721 QTVLHSSFEH VDYRVTGLFL AALEVIRRGQ WNFYRLENEH LNNAGKFRAV KTVPLPFREV 780
781 DEED 784
```

Figure 4.8 Amino acid sequence of AtPHO1;H1.

N-terminal SPX domains from AtPHO1;H1 of different lengths were selected for expression test in *E. coli* cells. AtPHO1;H1_1 containing residues 1-340 and AtPHO1;H1_2 containing residues 1-384 are highlighted in yellow and underlined, respectively.

Solubility screening of varying SPX expression constructs

Using the above selected SPX domains, expression constructs were made with different affinity tags by sub-cloning the SPX domain coding regions into pET28, pL53 and pGEX-2T expression vectors (Appendix 2 for plasmid maps). Since every protein is different with respect to the optimal conditions for recombinant protein production, different expression conditions, including combinations of a range of *E. coli* host strains (Miroux and Walker, 1996; Wagner *et al.*, 2008), culture medium, expression induction methods (Studier, 2005) and expression temperatures were tested on these SPX domain expression constructs to increase the solubility of over-expressed SPX proteins. In all, 5 different constructs for codon optimized AtSPX1 using His_6, His_8 and streptavidin binding peptide C terminal tags were constructed and tested. Two different truncations of AtPHO1;H1 with either C terminal His_6 tags or N terminal maltose binding protein tags were constructed and tested. Finally non-codon optimized AtSPX1 and AtSPX3 were constructed as N terminal GST fusions (by MSc student Dikani Salema). All constructs were verified by restriction digestion (Figure 4.9) and sequencing. Target protein expression and solubility was examined by initial dot blotting and subsequent western blotting with corresponding antibodies for each affinity tag (Table 4.1).
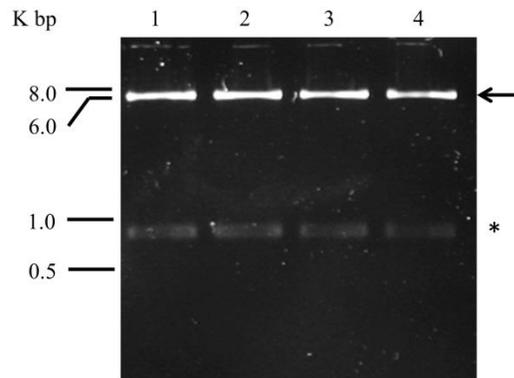
Figure 4.9 Restriction digestion of construct pET28-AtSPX1-His$_6$.

Codon usage optimized sequence coding AtSPX1 was synthesized and sub-cloned into pET28 vector using available *Nco*I/*Xho*I sites to generate a C-terminal His$_6$ fused construct. Presence of inserted *AtSPX1* sequence in successful transformants (1-4) was checked by double restriction digestion using enzyme *Nco*I and *Xho*I. Digested fragments were analyzed on 1% agarose gel. DNA fragment corresponding to the pET28 vector backbone (5233 bp) and the inserted AtSPX1 sequence (771 bp) are indicated by an arrow and *, respectively.

Table 4.1 Expression screen of the different SPX domain expression constructs.

SPX domain (residues)	Tag	Medium	Strains[¤]	Expression	Solubility
AtSPX1* (full-length)	His$_6$ [C] [a]	M9, LB, SB	BL21-gold, BL21-star, C41, C43, Lemo21	+	+
AtSPX1* (full-length)	His$_8$ [C] [a]	M9, LB, SB	BL21-gold, BL21-star, C41, C43	+	-
AtSPX1* (full-length))	SBP-His$_8$ [C] [a]	M9, LB, SB	BL21-gold, BL21-star, C41, C43	+	-
AtSPX1*-S1 (1-159)	His$_6$ [C] [a]	LB, SB	BL21-gold, BL21-star, C41, C43	+	-
AtSPX1*-S2 (1-181)	His$_6$ [C] [a]	LB, SB	BL21-gold, BL21-star, C41, C43	+	-
AtSPX1 (full-length)	GST [N]	LB, SB	BL21	+	+
AtSPX3 (full-length)	GST [N]	LB, SB	BL21	-	N/A
AtPHO1;H1_1 (1-340)	His$_6$ [C] [a]	M9, LB, SB	BL21-gold, BL21-star, C41, C43, Lemo21	+	-
AtPHO1;H1_1 (1-340)	His$_8$-MBP [N]	M9, LB, SB	BL21-gold, BL21-star, C41, C43	-	N/A
AtPHO1;H1_2 (1-384)	His$_6$ [C] [a]	M9, LB, SB	BL21-gold, BL21-star, C41, C43, Lemo21	+	-
AtPHO1;H1_2 (1-384)	His$_8$-MBP [N]	M9, LB, SB	BL21-gold, BL21-star, C41, C43	-	N/A

*: Codon optimized coding sequences
[¤]: All bacteria strains used are DE3 strains
[a]: Constructs using pET vectors

[C]: C-terminal tags
[N]: N-terminal tags
SBP: Streptavidin Binding Peptide
MBP: Maltose-Binding Protein
GST: Glutathione S-Transferase

M9: M9 Minimal Media
LB: Standard Lysogeny Broth
SB: Super Broth

The two N-terminal His$_8$-MBP fused expression constructs of AtPHO1;H1 SPX domains resulted in the detachment of N-terminal tags from the target SPX protein. A major band corresponding to the size of N-terminal His$_8$-MBP tag (41.7 kDa) was seen in the expression test of these two proteins using BL21-gold (DE3) strain, whereas no obvious over-expressed protein was seen at the expected size of His$_8$-MBP-AtPHO1;H1_1 (83.9 kDa) or His$_8$-MBP-AtPHO1;H1_2 (89.2 kDa) (Figure 4.10 A). Expression tests carried out at different time points during expression showed the detachment of N-terminal tags occurred during expression (Figure 4.10 B). Expression of these two constructs in other *E. coli* strains resulted in the same detachment of N-terminal tags. Due to the lack of C-terminal tag, the expression and solubility of the target SPX protein was rendered unmeasurable.
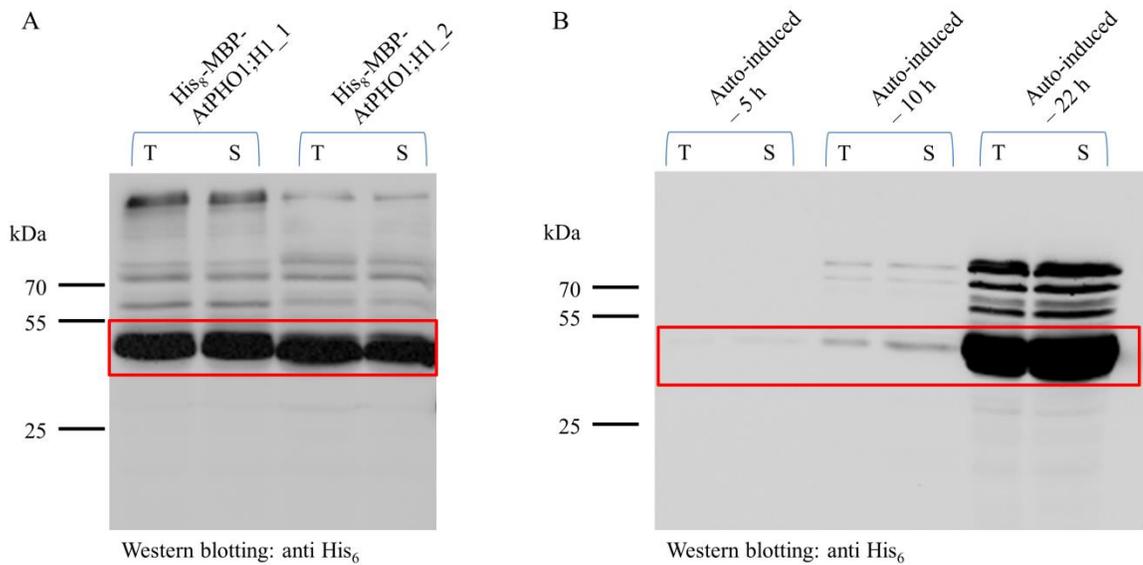
Figure 4.10 Expression test of His$_8$-MBP-AtPHO1;H1_1 and His$_8$-MBP-AtPHO1;H1_2 in BL21-gold (DE3) strain.

Cells over-expressing His$_8$-MBP-AtPHO1;H1_1 and His$_8$-MBP-AtPHO1;H1_2 were harvested after 22 h of expression (A) and cells over-expressing His$_8$-MBP-AtPHO1;H1_2 were harvested at different time points of expression (B). Total and soluble proteins were extracted using native cell lysis buffer (Chapter 2 section 2.1.9.1 and 2.2.5.1). An equivalent amount of proteins from total protein extract (T), soluble protein fraction (S) was separated on 12% SDS-PAGE gel and transferred to a nitrocellulose membrane for immunoblotting using anti-His$_6$ antibody. The detached N-terminal tags (41.7 kDa) are marked with red rectangles.

Nevertheless, initial expression screens showed that apart from these two constructs and N-terminal GST tagged AtSPX3, all the other versions of SPX domain from AtSPX1 and AtPHO1;H1 were able to be expressed in *E. coli* cells. However, most of the constructs were not expressed in a soluble form under any of the conditions tested (Table 4.1). Using Lemo21, a derivative strain from BL21 (DE3) that expresses target proteins at lower rates (Wagner *et al.*, 2008) did not improve the solubility of SPX domain. This suggests the possible accumulation of unfolded over-expressed proteins in the inclusion bodies. As it has been shown that active recombinant proteins can be generated by inclusion body solubilization followed by *in vitro* refolding (Rudolph and Lilie, 1996; Vallejo and Rinas, 2004), the insoluble AtSPX1-SBP-His$_8$ and AtSPX1-His$_8$ proteins were solubilized from inclusion bodies using 6 M guanidine hydrochloride (GndCl) and purified under a denatured condition via IMAC (Figure 4.11). Purified denatured AtSPX1-SBP-His$_8$ and AtSPX1-His$_8$ proteins were refolded by dialysis into native buffer that contains 400 mM L-arginine (Chapter 2 section 2.2.5.2).

The denaturing extraction and purification results showed that insoluble AtSPX1-SBP-His$_8$ and AtSPX1-His$_8$ could be solubilized with 6 M GndCl and extracted from the inclusion bodies (Figure 4.11 A). IMAC purification under denatured condition showed a relatively high recovery of denatured target proteins in the elution fractions (Figure 4.11 A). High purity of refolded AtSPX1 proteins was obtained after denaturant removal (Figure 4.11 B). Only one major contaminant was observed below 10 kDa and western blotting result indicated this is likely to be the C-terminal tags as it was recognized by the anti-His$_6$ antibody (Figure 4.11 B).
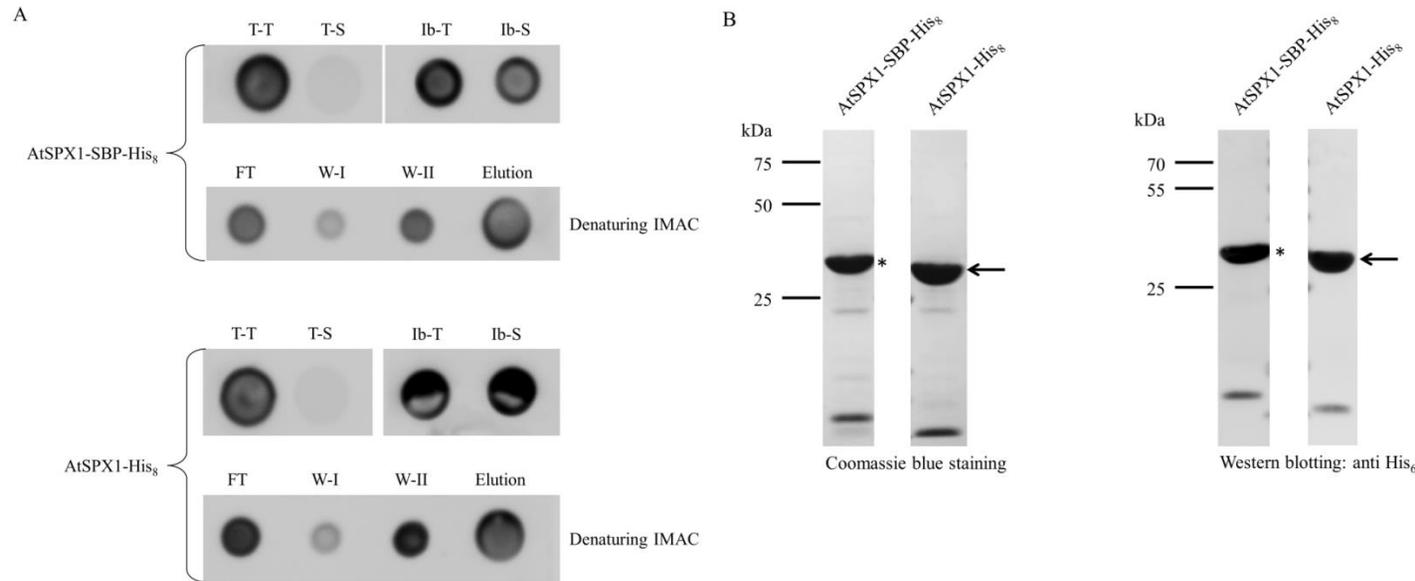
Figure 4.11 Denaturing purification and refolding of AtSPX1-SBP-His$_8$ and AtSPX1-His$_8$ proteins.

Cells over-expressing recombinant AtSPX1-SBP-His$_8$ and AtSPX1-His$_8$ were lysed and inclusion bodies were washed and solubilized in denaturing buffer containing 6 M GndCl (Chapter 2 section 2.2.5.2). AtSPX1-SBP-His$_8$ and AtSPX1-His$_8$ were purified via IMAC using Ni-NTA resin under denatured conditions. Eluted proteins from IMAC were dialyzed to remove GndCl and imidazole. Cell lysis and denaturing purification fractions were checked by dot blotting with anti-His$_6$ antibody (A) and an equivalent amount of dialyzed proteins were separated on 12% SDS-PAGE gel and stained using Coomassie blue (B left panel). Presence of His$_8$-tag proteins was detected by western blotting with anti-His6 antibody (B right panel). (T-T: Total cell lysate; T-S: Supernatant cell lysate; Ib-T: Total 6 M GndCl solubilized inclusion body; Ib-S: Soluble fraction from 6 M GndCl solubilized inclusion body; FT: Unbound flow-through from denaturing IMAC; W-I / W-II: wash fractions from denaturing IMAC; Elution: elution fractions from denaturing IMAC.) The predicted mobility of AtSPX1-SBP-His$_8$ (37.5 kDa) and AtSPX1-His$_8$ (32.9 kDa) was indicated by an asterisk (*) and an arrow (←), respectively.

However, although high purity of refolded AtSPX1-SBP-His$_8$ and AtSPX1-His$_8$ was obtained and a high yield of purified proteins (around 40 mg protein per liter cell culture) was estimated by light absorbance at 280 nm, circular dichroism (CD) spectrum on the refolded proteins failed to show any feature of alpha-helical structure (Wild *et al.*, 2016), and *in vitro* protein interaction assay with MBP-AtdPHR1 (Chapter 5 section 5.2.2) did not show any evidence of binding to AtdPHR1 as seen from the natively purified GST-AtSPX1 (Chapter 5 section 5.2.3). Therefore, the AtSPX1 protein obtained under the denatured condition is likely to be incorrectly refolded and the investigation using constructs giving non-soluble proteins was not followed up. The subsequent expression and purification work was focused on two constructs AtSPX1-His$_6$ and GST-AtSPX1 that generated soluble proteins.

Dot blotting showed soluble AtSPX1 protein was obtained from the construct AtSPX1-His$_6$ when expressed at 20 °C in SB auto-induction medium using BL21-gold (DE3) strain and GST-AtSPX1 expressed at 18 °C in LB IPTG-induction medium using BL21 (DE3) strain (Figure 4.12 A). Subsequent western blotting also confirmed the presence of target proteins at the expected sizes using corresponding antibodies (Figure 4.12 B). Dot blotting and western blotting results also showed higher protein solubility from AtSPX1-His$_6$ compared to GST-AtSPX1 when extracted with native buffers (Chapter 2 section 2.1.9.1 and 2.2.5.1) although a major protein degradation product between 15 – 20 kDa was seen from AtSPX1-His$_6$ (Figure 4.12 B). These two constructs and the corresponding expression conditions were used for the initial affinity purification trials on soluble AtSPX1.
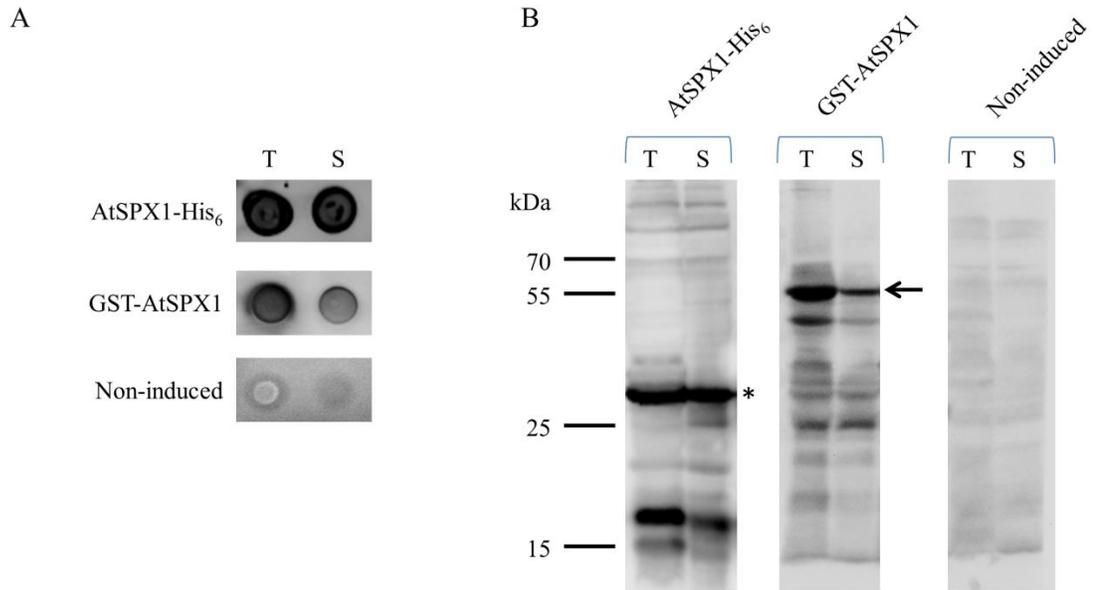
Figure 4.12 Expression test on C-terminal His$_6$ fused AtSPX1 and N-terminal GST fused AtSPX1.

C-terminal His$_6$ fused AtSPX1 and N-terminal GST fused AtSPX1 when expressed at 20 °C in SB auto-induction medium using BL21-gold (DE3) strain and at 18 °C in LB IPTG-induction medium using BL21 (DE3) strain, respectively. Presence of target protein AtSPX1-His$_6$ and GST-AtSPX1 in total cell lysate (T) and soluble lysate (S) were detected by anti-His$_6$ antibody and anti-GST antibody, respectively. A. Dot blotting of AtSPX1-His$_6$ and GST-AtSPX1 expression test; B. Western blotting of AtSPX1-His$_6$ and GST-AtSPX1 expression test, AtSPX1-His$_6$ (Theoretical Mw 31.2 kDa) and GST-AtSPX1 (Theoretical Mw 56.4 kDa) are indicated by * and an arrow, respectively.

Initial affinity purification trials of selected constructs

Initial AtSPX1 purification using affinity chromatography was carried out using the C-terminal His_6 fused construct, as higher solubility was seen from the expression test (Figure 4.12). After the target protein was expressed under the conditions described above, cells were lysed by disruption in native protein extraction buffers containing protease inhibitor (Chapter 2 section 2.2.5.2) and recombinant AtSPX1-His_6 was purified via IMAC using Ni-NTA resin (Figure 4.13).

The IMAC purification results shown in Figure 4.13 indicate that the over-expressed recombinant AtSPX1 was able to be largely extracted from host cells. High protein solubility was also maintained as similar western blotting signal intensity was seen from soluble lysate fraction compared to the total cell lysate (Figure 4.13 B). However, the yield of purified protein was strongly reduced by extremely low binding efficiency as most of the soluble fusion protein was detected in the unbound flow-through fraction after column binding (Figure 4.13 B). Although western blotting with monoclonal anti-His_6 antibody showed most of the column bound target protein was eluted with 40-60 mM imidazole, Coomassie blue stained gel showed the amount of target protein recovered in the elution fractions is barely detectable and much lower than the other co-purified contaminants (Figure 4.13 A). The major degradation product seen in the previous expression test is also detected in western blotting from most of the purification fractions (indicated by * in Figure 4.13 B), suggesting the N-terminal regions of this AtSPX1 protein is not stable and degradation occurs even in the presence of protease inhibitors. Other contaminating bands of native proteins from *E. coli* host cells are also observed, including a strong band around 50 kDa that possibly represent the native metal-binding protein ArgE (Bolanos-Garcia and Davies, 2006).
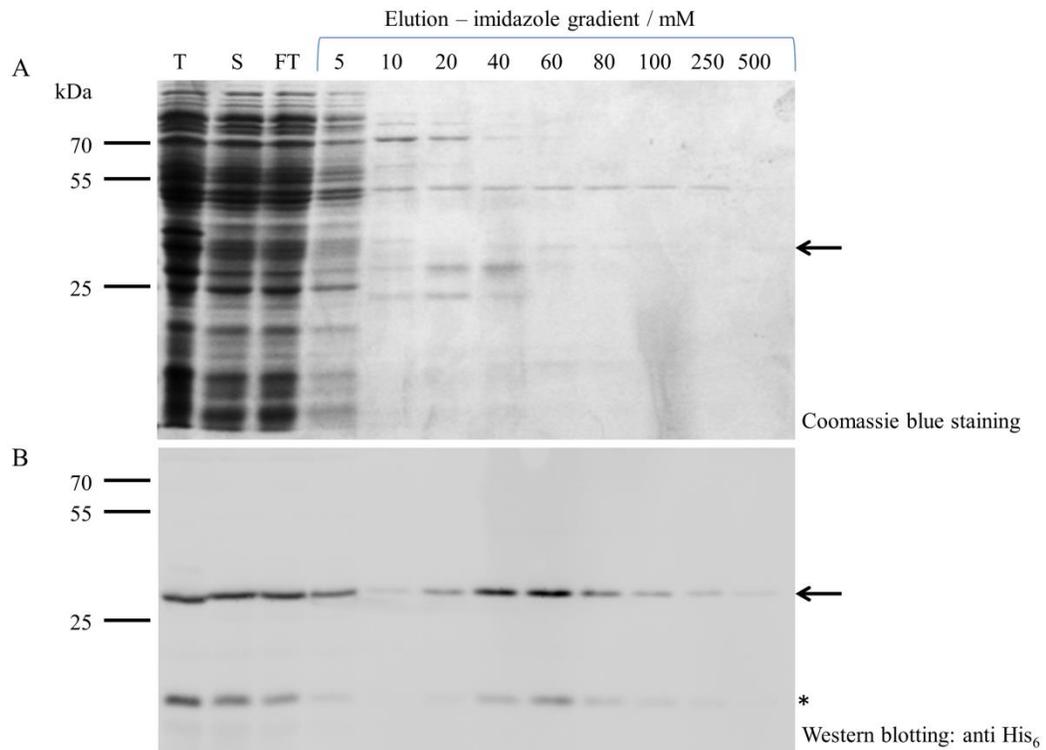
Figure 4.13 Immobilized metal affinity chromatography (IMAC) purification of AtSPX1-His$_6$.

IMAC purification of AtSPX1-His$_6$ was carried out using Ni-NTA resin on a gravity flow column. After cell lysis, soluble extraction fraction was incubated with resin at 4°C for 1 h before bound proteins were eluted with a gradient concentration of imidazole. Equivalent amount of proteins from total cell lysate (T), soluble cell lysate (S) unbound flow-through (FT) and elution fractions was separated on 12% SDS-PAGE gel and stained using Coomassie blue (A). Equal amount of proteins were also transferred to a nitrocellulose membrane after SDS-PAGE for immunoblotting using anti-His$_6$ antibody (B). AtSPX1-His$_6$ (Theoretical Mw 31.2 kDa) is indicated by arrows and the major contaminant band from western blotting is indicated by an asterisk (*).

Varying protein incubation time with the resin had little effect on the binding efficiency, as most of the target protein is still seen in the unbound flow-through fraction even after 3 hours incubation as detected by western blotting using anti-His$_6$ antibody (Figure 4.14), suggesting a generally poor accessibility of the C-terminal His$_6$ affinity tag. This also corresponds with most of the target protein being eluted with a low imidazole concentration at 40 mM (Figure 4.13), which is likely to be the result of weak protein binding to the resin.

Although the C-terminal His$_6$ fused AtSPX1 showed high solubility when over-expressed in *E. coli* cells, the low column binding efficiency during first step IMAC purification prevented the subsequent purification of sufficient quantity of protein for functional studies. An extended His$_8$ tag with or without an upstream Streptavidin Binding Peptide (SBP) tag was fused to the C-terminal of AtSPX1 in the attempt to increase the protein binding efficiency to Ni column. However, these constructs failed to produce soluble proteins in the expression screen (Table 4.1).
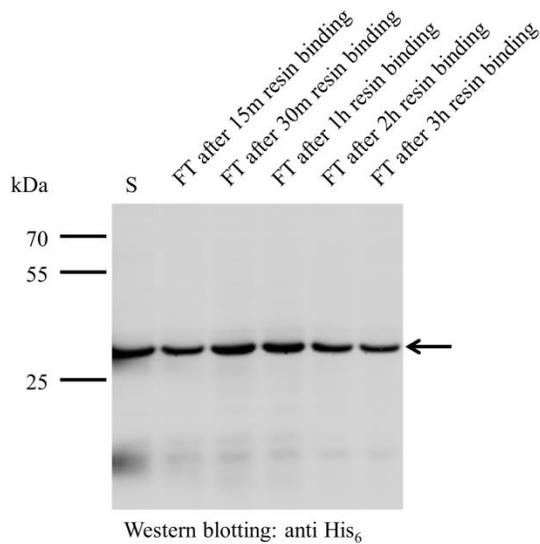
Figure 4.14 Effect of incubation time on AtSPX1-His$_6$ binding to Ni-NTA resin.

Soluble lysate from AtSPX1-His$_6$ over-expressing cells was incubated with Ni-NTA resin at 4°C for various time courses to examine the effect of incubation time on AtSPX1-His$_6$ resin binding. Equivalent amount of protein from soluble cell lysate (S) and unbound flow-through fractions (FT) was separated on 12% SDS-PAGE gel and target protein was detected by western blotting using anti-His$_6$ antibody. AtSPX1-His$_6$ (Theoretical Mw 31.2 kDa) is indicated by an arrow.

Affinity chromatography was then carried out on the N-terminal GST fused AtSPX1 using glutathione resin (Figure 4.15). Due to the slow binding kinetics between GST and glutathione, the initial purification was performed at room temperature for 30 min as recommended by the manufacturer's instructions (GE Healthcare) to increase the column binding efficiency.

The purification result showed the soluble GST-AtSPX1 was able to be recovered after affinity chromatography, although degradation products were observed in all the elution fractions, with a major degradation product of ~25 kDa likely corresponding to the N-terminal GST tag since this fragment cross reacts with the anti-GST antibody (Figure 4.15). Other high molecular weight contaminants were also co-eluted (Figure 4.15 A), suggesting non-specific binding of endogenous proteins to the glutathione column, including a ~70 kDa band likely corresponding to the host cell Hsp70s protein that assists the protein folding processes (Mayer and Bukau, 2005). In order to reduce the AtSPX1 degradation and prevent co-purification of contaminating proteins, affinity purification of GST-AtSPX1 was tested at 4°C with prolonged column binding and protein elution time (Figure 4.16).

SDS-PAGE analysis on elution fractions using equivalent amount of proteins showed that although similar amounts of protein degradation products are still co-purified, the recovery rate of target protein GST-AtSPX1 is much higher at 4°C (Figure 4.16). Less contamination of high molecular weight proteins was also detected (Figure 4.16). This suggests less protein aggregation or on column precipitation occurred at lower temperature and that the target AtSPX1 protein might be more stable under 4°C compared to room temperature. Being expressed in a soluble form, the N-terminal GST-tagged AtSPX1 construct demonstrated a relatively high recovery and purity of target protein after first step affinity purification, and is therefore chosen for further functional and biochemical study in the next chapter.
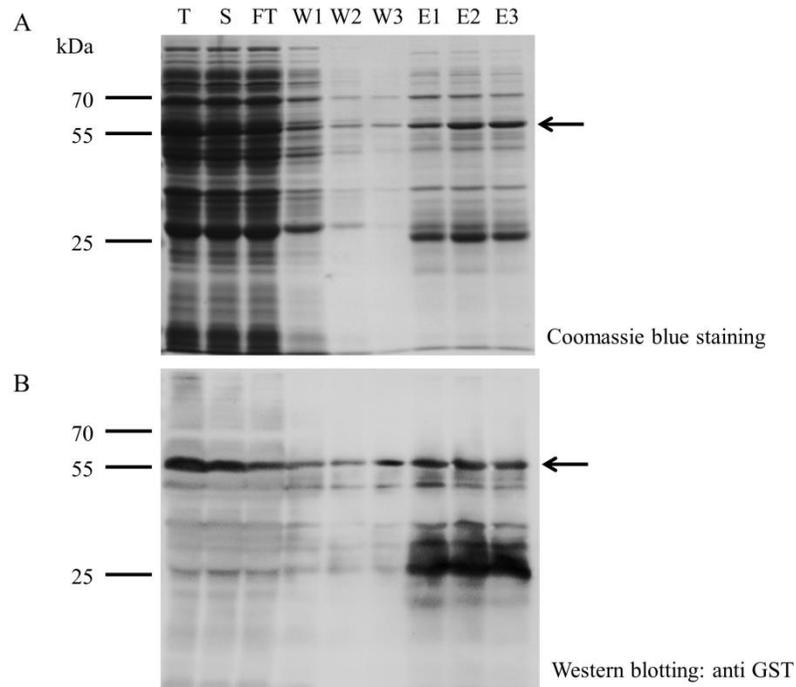
Figure 4.15 Affinity chromatography purification of GST-AtSPX1.

Affinity chromatography of GST-AtSPX1 was carried out using glutathione resin on a gravity flow column. After cell lysis, soluble cell lysate was incubated with resin at room temperature for 30 min before bound proteins were eluted with 10 mM reduced glutathione. Equivalent amount of protein from total cell lysate (T), soluble cell lysate (S) unbound flow-through (FT), column washing fractions (W1-3) and elution fractions (E1-3) was separated on 12% SDS-PAGE gel and stained using Coomassie blue (A). Equal amount of protein were also transferred to a nitrocellulose membrane after SDS-PAGE for immunoblotting using anti-GST antibody (B). GST-AtSPX1 (Theoretical Mw 56.4 kDa) is indicated by arrows.

kDa

Affinity Purification at room temperature
Affinigy purification at 4°C

70
55
25

Coomassie blue staining
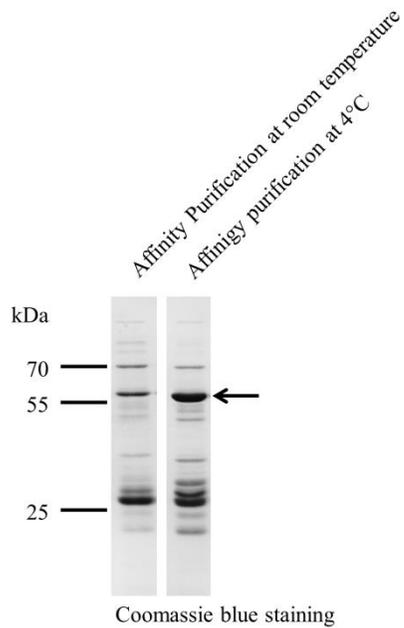
Figure 4.16 GST-AtSPX1 affinity chromatography at different temperatures.

Affinity chromatography of GST-AtSPX1 was carried out using glutathione resin at either room temperature or 4°C. After cell lysis, soluble cell lysate was incubated with resin either at room temperature for 30 min or at 4°C for 2 h. Target proteins were eluted with 10 mM reduced glutathione by incubating in eluting buffer at room temperature for 5 min or at 4°C for 30 min. Equivalent amount of eluted proteins from room temperature and 4°C purification was separated on 12% SDS-PAGE gel and stained using Coomassie blue. GST-AtSPX1 (Theoretical Mw 56.4 kDa) is indicated by an arrow.

4.3 Discussion

While prokaryotic cells have been shown to be capable of expressing functional eukaryotic proteins, there is still the concern that the lack of certain protein folding and post-translational modification machinery in the prokaryotic expression system may lead to incorrectly folded or inactive recombinant eukaryotic proteins (Khow and Suntrarachun, 2012). Since the highly conserved SPX protein domain is shown to be present in many different eukaryotic species (Secco *et al.*, 2012a) but no prokaryotic protein has been identified possessing this domain, the expression of AtSPX1 was carried out using both a plant expression system and a bacterial expression system. Although using the binary pEAQ-*HT* vector was shown to achieve high levels of GFP protein expression at ~1.5 g/kg fresh tissue in agro-infiltrated tobacco leaves (Sainsbury *et al.*, 2009), the yield of recombinant AtSPX1 was much lower when expressed using the same system. The expression of AtSPX1 using a C-terminal His$_6$ tag resulted in an optimal protein yield of 35 mg/kg leaf tissue, whereas the N-terminal His$_6$ tag fused version failed to show any detectable expression (Figure 4.3). However, it is notable that in Sainsbury *et al.*, 2009, when using the pEAQ-*HT* vector for expression, fusion to a C-terminal His$_6$ tag is reported to reduce the GFP protein production by two thirds and more pronounced expression reduction was seen from the N-terminal His$_6$ tag fusion. Therefore, the expression of AtSPX1 in tobacco is very likely affected by the fused His$_6$ tags. Nevertheless, the agro-infiltrated tobacco expression system still showed reasonable expression of C-terminal His$_6$ fused AtSPX1. Comparable levels of full-length recombinant Human Papillomavirus 8 L1 protein (60 mg/ kg leaf tissue) as well as chimeric Human papillomavirus 16 L1 protein (45-120 mg/ kg leaf tissue) being obtained using this virus-derived pEAQ-*HT* vector has also been reported (Matić *et al.*, 2011; Matić *et al.*, 2012). Unfortunately, as the soluble AtSPX1 protein showed very low recovery after purification from plant tissue, possibly due to aggregation or on column precipitation caused by low stability of C-terminal His$_6$ fused AtSPX1, this system was not chosen for obtaining AtSPX1 for further studies.

When expressing heterologous plant proteins in *E. coli* host cells, the total expression may be compromised due to the different codon usage of plant genes compared to the bacterial host (Khow and Suntrarachun, 2012; Grote *et al.*, 2005). In order to increase the protein

expression, the gene coding heterologous proteins can be optimized to the preferred codon usage in *E. coli* or codon bias-adjusted *E. coli* strains can be used for the over-expression (Burgess-Brown *et al.*, 2008). In this study, both codon optimized and wild type *Arabidopsis* SPX domain coding genes were used for expression in *E. coli* host strains with and without rare codon recognizing tRNAs. As the expression test results show that both the codon optimized and wild type SPX domains coding genes showed reasonable protein expression, it suggests the codon optimization is not critical for the over-expression of *Arabidopsis* SPX proteins. Similarly, as *AtSPX3* failed to express using the native coding gene in this study, other studies trying to over-express *AtSPX3* in codon bias-adjusted *E. coli* host cells using codon optimized sequences didn't show protein expression either (Wild *et al.*, 2016), further demonstrating codon optimization does not improve the expression of SPX domains.

On the other hand, as many SPX constructs resulted in insoluble proteins, it argues the codon optimization of coding genes or including of rare codon tRNAs in host cells might have caused the saturation of protein folding machinery in the *E. coli* system due to high level of over-expression and therefore lead to protein aggregation and insolubility, as observed in previous studies (Rosano and Ceccarelli, 2009). As higher solubility of recombinant AtSPX1-His$_6$ and GST-AtSPX1 was also observed when proteins were expressed at lower temperatures, this indicates a lower translation rate might be beneficial to protein solubility. However, reducing initial expression rate using L-rhamnose controlled Lemo21 expression strain (Wagner *et al.*, 2008) did not improve the solubility of SPX domains (data not shown), suggesting the solubility of this protein domain is not solely dependent on the expression rate but might also be affected by other factors such as affinity tags or versions of SPX domains used for expression. Previous protein over-expression investigations in *E. coli* host cells have shown adding affinity tags such as N-terminal MBP significantly accelerate the folding and solubility of its fusion partners (Kapust and Waugh, 1999; Raran-Kurussi *et al.*, 2015), although in this study the MBP tag was detached from the fusion AtPHO1;H1 proteins before purification. Many other commonly used affinity tags such as a His-tag, however, do not contribute to the solubility of fused proteins but only promote the purification of fusion proteins (Waugh, 2005). This might explain while the C-terminal His$_6$ tagged AtSPX1 showed high solubility, increasing the length of the His tag rendered the construct insoluble.

Although the SPX domain itself is a hydrophilic domain, the expression domain boundary choice at the C-termini could be vital in determining the solubility (Dyson, 2010). Due to the lack of SPX domain structures when the *Arabidopsis* SPX constructs were made, the SPX domain coding regions for AtPHO1;H1_1 and AtPHO1;H1_2 were obtained by amino acid sequence alignment and each domain boundary was chosen before a poorly conserved residue. As the accuracy of domain boundaries predicted by alignment might be strongly affected by sequence variation from different species, this may induce the risk of interrupting the SPX sub-domain boundaries and subsequently reduce the solubility of chosen SPX domains from AtPHO1;H1. Recent structural studies on another PHO1 family member AtPHO1 showed a soluble SPX domain was obtained using a construct containing residue 1-386 from AtPHO1 (Wild *et al.*, 2016) (Figure 4.17). This corresponds to residues 1-387 of AtPHO1;H1, which is 3 amino acids longer at the C-terminus than construct AtPHO1;H1_2 (residues 1-384), suggesting the insolubility of this construct is very likely caused by an interrupted soluble domain boundary (Figure 4.17).

Study on the available SPX domain structures showed this conserved protein domain, although adopting varying organization in different proteins, contains six α helices as its main feature (Wild *et al.*, 2016) (Figure 4.17). Assignment of these main secondary structures (Wild *et al.*, 2016) indicates the C-terminal boundaries chosen for AtPHO1;H1_1 (residue 1-340) and AtSPX1-S1 (residue 1-159) fall between helices α5 and α6 at the C-termini of SPX domain, whereas the domain boundary chosen for AtSPX1-S2 falls in the middle of the α6 helix (Figure 4.17), which extends into the predicted disordered C-terminal region of AtSPX1 (Figure 4.7). As these two C-terminal helices α5 and α6 together are found to form a 3-helix bundle with the core helices α3 and α4 (Wild *et al.*, 2016), separating these two helices or removing the AtSPX1 C-terminal disordered regions could result in unstable SPX domain C-termini and subsequent insolubility of recombinant proteins as seen from the expression result in this chapter (Table 4.1).

```
        ⊢ α1 ⊣              ⊢―――――――――― α2 ――――――――――⊣                                    ⊢――――――――― α3 ――――――――――
           10        20        30        40        50        60        70        80        90       100       110       120       130       140
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AtSPX1           -MKFGKSLSNQIEQTLPEWQDKFLSYKELKK------------------------------RLKLIGSKTADRP-----------------VKRLRLDEFSVGISKEEINFIQLLEDELEKFNNFFVEKEEEYIIR---L
AtPHO1;H1        MVKFTKQFEG---QLVPEWKDAFVDYSQLKKDLKKIHLFTNGVEKKHTETSLIKTVKSSLGRLSIFGNKGREQSRVIQVHKKLASSGSNNDVYETELLEKIADDTDAAKEFFACLDMQLNKVNQFYKTKEKEFLERGECL
Clustal Consensus :** *.:..   * :***:* *:.*.:***                **.::*.* ::.              * . .* *  .. :.   :*: *: :*:*.*:*: **:*:: *    *

                                                                                                   ⊢――――――――――――――――――――――――― α4 ――――――――――――――――――――――――――
                       150       160       170       180       190       200       210       220       230       240       250       260       270       280
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AtSPX1           KEFRDRIAKAKDSMEKMIK------------------------------------------------------------------------------------------------------------------------
AtPHO1;H1        KKQMDILIELKDAFKQKQANGESTQESKEDDSISCTISCEYDSVRGRTEEMQLQVSCLDNLEDNGEEALESLGSEEPIKANNEDSKLTTVSSRVFSCQGKNVKIKIPLTNPSRTFSAISYLINQSSSKKNGPDGGNKLQI
Clustal Consensus *:  * : : **::::

                                                                            ⊢ α5 ⊣        ⊢――――――――――――――――――――― α6 ―――――――――――――――――――――⊣
                       290       300       310       320       330       340       350       360       370       380       390       400       410       420
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AtSPX1           ----------IRKEIVDFHGEMVLLENYSALNYTGLVKILKKYDKRTGDLMRLPFIQKVLQQPFYTTDLLFKLVKESEAMLDQIFPANET------------------------------ESEIIQAELS------
AtPHO1;H1        SKKKLSHAEKMIKGALTELFKGLNYLKTYRNLNILAFMNILKKFDKVTGKQILPIYLKVVESSYFNISDKVMILSDEVEEWFIKHLAGENRRKAMKYLKPHHRKESHSVTFFIGLFTGCFVALLAGYIIVAHLTGMYRQH
Clustal Consensus       *: :.::. :  *:.* **  .::*****:** **. :   ::: * .. *  :* :: * .* * : : :..::                 . ** *.*:

                       430       440       450       460       470       480       490       500       510       520       530       540       550       560
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AtSPX1           ------------------------------------EHKFMESLHMKSTIAALRVLKEIRSGSSTVS--------------------------------------------------------VFSLPPLQLN
AtPHO1;H1        SANTFYMETAYPVLSMFGLLFLHLFLYGCNIFMWRKARINYSFIFELGSKNELKYRDVFLICTASMSAIAGVMFVHLSLLEKGYSFRQVQVIPGLLLLGFLLILICPLNIFYKSSRYRLISVIRNIVFSPLYKVVMLDFF
Clustal Consensus                   ::.*: .* *. :   *:   :. *:::                                                                  ::.: *::

                       570       580       590       600       610       620       630       640       650       660       670       680       690       700
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AtSPX1           GLDETWKKIPLLEQEAK-----------------------------------------------------------------------------------------------------------------------
AtPHO1;H1        MADQLCSQVPMLRNLEYIACYYITGSYATQDYEYCMRVKYYRDLAYAVSFLPYYWRAMQCARRWFDEGETSHLVNLGKYVSAMLAAGTKVAYEKERSLGWLCLVVAMSSVATIYQLYWDFVKDWGLLQHNSNNPWLRNQL
Clustal Consensus   *:  .::*:*.:

                       710       720       730       740       750       760       770       780
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AtSPX1           --------------------------------------------------------------------------
AtPHO1;H1        MLRQKSIYYFSMVLNLVLRLAWLQTVLHSSFEHVDYRVTGLFLAALEVIRRGQWNFYRLENEHLNNAGKFRAVKTVPLPFREVDEED
Clustal Consensus
```

Figure 4.17 Sequence alignment of AtSPX1 and AtPHO1;H1 including a secondary structure assignment from SPX[ScVtc4] (Wild et al., 2016).

The AtSPX1 and AtPHO1;H1 protein sequences were aligned using Clustal X (Higgins et al., 1996; Thompson et al., 1997). A secondary structure of α-helices assignment from SPX[ScVtc4] (Wild et al., 2016) was marked on top of the corresponding sequence regions. SPX domains used for constructs AtSPX1-S1 and AtSPX1-S2 are highlighted in yellow and single underlined, respectively. SPX domains used for constructs AtPHO1;H1_1 and AtPHO1;H1_2 are highlighted in turquoise and double underlined, respectively.

Due to the N-terminal region of the SPX domain being highly conserved, its natural state was initially preserved by adding affinity tags to the C-terminus of most AtSPX1 expression constructs (Table 4.1). However, although the C-terminal His$_6$ fused AtSPX1 expression constructs showed very high solubility, a low protein recovery after IMAC purification was observed due to poor binding efficiency. As the α6 helix at the C terminus of other homologue SPX domains is shown to adopt different orientations in different structures (Wild *et al.*, 2016), it might have caused the C-terminal His$_6$ tag from AtSPX1 to be largely buried in the secondary structure and therefore not accessible for interacting with the metal ions during IMAC purification. Although Wild *et al.* were able to express and purify enough protein for structural studies on the fungal and mammalian SPX domains, they have demonstrated high instability of the SPX domain N-terminus especially without substrate binding (Wild *et al.*, 2016). The N-terminal GST fused construct, when expressed at lower temperature, produced soluble recombinant protein that is sufficient for subsequent biochemical studies (discussed in chapter 5). Although the same conditions for expressing GST-AtSPX1 recombinant protein in *E. coli* host cells were also used in other studies to produce soluble AtSPX1 for functional assays (Puga *et al.*, 2014), the recombinant AtSPX1 obtained from *E. coli* expression system is not stable enough for structural investigations, as degradation was seen even with the presence of protease inhibitors (Figure 4.15 and 4.16). Other prokaryotic expression attempts also failed to produce stable recombinant plant SPX exclusive family proteins for structural studies (Wild *et al.*, 2016). Since binding to substrates such as inositol polyphosphate (InsPs) is known to promote the interaction between plant SPX proteins and their *in vivo* partners (Puga *et al.*, 2014; Wang *et al.*, 2014) and to potentially stabilize SPX domain N-terminal helices through additional hydrogen bonds (Wild *et al.*, 2016), it is likely that the *E. coli* expressed SPX proteins may exhibit low stability due to the lack of natural interacting partners. This further suggests yeast (*Saccharomyces cerevisiae*) cells, which encode 10 SPX domain-containing proteins and their interacting partner Spl2 (Hürlimann *et al.*, 2009), could be a potentially good system for plant SPX protein expression.

Chapter 5   Purification and functional analysis of AtSPX1 and its interacting partner AtPHR1

5.1   Introduction

Arabidopsis SPX1 has been shown to interact with and inhibit the central transcription factor in phosphate response, AtPHR1, in a Pi-dependent manner (Puga *et al.*, 2014). However, the molecular mechanism of the interaction is not clear. As the SPX protein domain was recently shown to physically interact with phosphate (Wild *et al.*, 2016), it is very likely that AtSPX1 is acting as a direct Pi binding protein. In this chapter, the capacity of AtSPX1 for binding Pi was tested using MicroScale Thermophoresis (MST) and the mechanism of AtSPX1-AtPHR1 interaction was investigated using Surface Plasmon Resonance (SPR). The GST-AtSPX1 construct obtained from Chapter 4 was used for the AtSPX1 purification, and a truncated version of AtPHR1 which lacks the transcription activation domain and is fused to MBP, termed MBP-AtdPHR1 was produced for the functional assays.

Two different AtSPX1-AtPHR1 interaction models can be proposed; either AtSPX1 displaces AtPHR1 from associating DNA molecules and thus turns off the transcription of PSI genes [as suggested by (Puga *et al.*, 2014) and (Wang *et al.*, 2014)], or AtSPX1 prevents AtPHR1 from binding to DNA and therefore modulates the PSI gene expression (as proposed and demonstrated here). The experiments in this chapter were designed to test a) whether SPX1 directly binds phosphate and b) to discriminate between these two models of regulation.

5.2 Results

5.2.1 Purification of GST-AtSPX1

Affinity purification of GST-AtSPX1

Previous work on the N-terminal GST tagged AtSPX1construct (Chapter 4) showed it to be more soluble when expressed at a lower temperature. Therefore the BL21 (DE3) cells expressing GST-AtSPX1 were grown at 37°C to reach the exponential growth phase and protein expression was carried out at 18°C. A lower concentration of isopropyl-β-D-thiogalactopyranoside (IPTG) was used (0.1 mM) with the aim of a slower overexpression rate permitting a greater percentage of correctly folded recombinant protein, ensuring a maximal yield of soluble AtSPX1. Prolonged cell growth time (~20 h) after induction was adopted to compensate the lower cell growth rate at the lower temperature. Cells expressing GST-AtSPX1 were lysed by cell disruption in the presence of protease inhibitors and recombinant GST-AtSPX1 was purified via affinity chromatography using glutathione sepharose resin (Figure 5.1).

This purification procedure typically had a yield of ~0.5 mg protein per liter of expression culture. Given the fact that quite a few contaminants are seen in the affinity purified GST-AtSPX1, it is difficult to estimate the yield of full length recombinant GST-AtSPX1 on its own. However, the western blotting result shows most of the contaminating proteins carry the GST-tag, apart from the band at ~70 kDa. This band is very likely representing Hsp70s from the host cells (Mayer and Bukau, 2005). This suggests GST-AtSPX1 may need molecular chaperones for correct folding and the stability of AtSPX1 might be low, and protein degradation is likely to occur even in the presence of protease inhibitors.
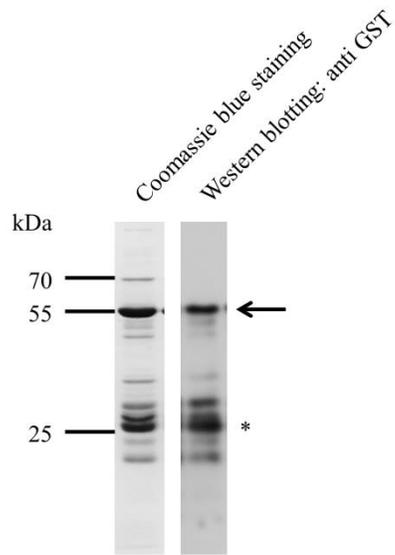
Figure 5.1 SDS-PAGE analysis of affinity chromatography purified GST-AtSPX1.

Affinity chromatography of GST-AtSPX1 was carried out using glutathione resin on a gravity flow column. Purified protein from the elution fraction was separated on 12% SDS-PAGE gel and stained using Coomassie blue (left panel). An equal amount of proteins were also transferred to a nitrocellulose membrane after SDS-PAGE for immunoblotting using anti-GST antibody (right panel). GST-AtSPX1 (Theoretical Mw 56.4 kDa) is indicated by an arrow. Many lower weight protein species such as degraded products are also lit up by the same antibody. Position of GST tag is marked by *.

Ion-Exchange chromatography purification of GST-AtSPX1

To further clean up the affinity purified GST-AtSPX1, a second purification step of ion-exchange chromatography was carried out using a HiTrap Q-HP 1 mL column with a gradient from 50 mM to 1 M NaCl in 10 mM HEPES pH 7.5 using an ÄKTA Explorer (Figure 5.2).

The addition of ion-exchange chromatography purification successfully removed most of the contamination proteins. However, the yield of GST-AtSPX1 was much lower after the second purification step (~0.15 mg protein per liter of expression culture), possibly due to protein aggregation and precipitation as high column pressure was generated during the purification. This again demonstrates the low stability of AtSPX1.

Coomassie blue staining
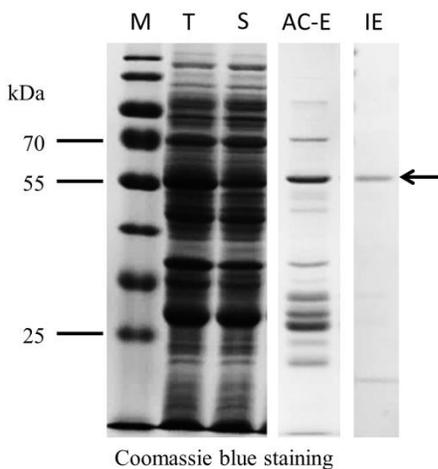
Figure 5.2 SDS-PAGE analysis of ion-exchange chromatography purified GST-AtSPX1.

GST-AtSPX1 was first enriched from crude cell lysate by affinity chromatography using glutathione resin before ion-exchange chromatography was carried out on the affinity chromatography eluates. An equivalent amount of proteins from total cell lysate (T), soluble cell lysate (S), affinity chromatography elution fraction (AC-E) and ion-exchange chromatography peak fraction (IE) was separated on 12% SDS-PAGE gel and stained using Coomassie blue. GST-AtSPX1 (Theoretical Mw 56.4 kDa) is indicated by an arrow. (M: Marker)

GST tag removal

The GST-AtSPX1 construct used contains a thrombin cleavage site between the GST and AtSPX1 domains. In order to get untagged AtSPX1, thrombin cleavage was attempted to remove the GST tag (Figure 5.3). The affinity purified GST-AtSPX1 protein was first exchanged into buffer optimal for thrombin cleavage (50 mM Tris-HCl pH 8.0, 10 mM $CaCl_2$) and then incubated with covalently bound thrombin-agarose resin and samples were taken after 1 h, 2 h and 4 h of cleavage. An equivalent amount of each cleavage fraction was loaded on SDS-PAGE gel for analysis (Figure 5.4 A). Cleaved GST and other GST-tagged un-cleaved proteins were removed by an additional affinity chromatography using glutathione sepharose resin (Figure 5.4 B).

As shown in Figure 5.4 A, thrombin protease efficiently removed the N-terminal GST tag after 1 h incubation and a protein band at ~26 kDa corresponding to GST was observed throughout thrombin cleavage. On the contrary, no accumulation of protein band at 30.2 kDa was observed, indicating the AtSPX1 protein was very likely precipitating or aggregating during the cleavage. Similarly, no un-tagged AtSPX1 protein was detected in the flow-through fraction of the following glutathione affinity chromatography (Figure 5.4 B), suggesting the AtSPX1 protein is very unstable on its own.

```
  1 MSPILGYWKI KGLVQPTRLL LEYLEEKYEE HLYERDEGDK WRNKKFELGL EFPNLPYYID 60
 61 GDVKLTQSMA IIRYIADKHN MLGGCPKERA EISMLEGAVL DIRYGVSRIA YSKDFETLKV 120
121 DFLSKLPEML KMFEDRLCHK TYLNGDHVTH PDFMLYDALD VVLYMDPMCL DAFPKLVCFK 180
181 KRIEAIPQID KYLKSSKYIA WPLQGWQATF GGGDHPPKSD LVPRGSMKFG KSLSNQIEQT 240
241 LPEWQDKFLS YKELKKRLKL IGSKTADRPV KRLRLDEFSV GISKEEINFI QLLEDELEKF 300
301 NNFFVEKEEE YIIRLKEFRD RIAKAKDSME KMIKIRKEIV DFHGEMVLLE NYSALNYTGL 360
361 VKILKKYDKR TGDLMRLPFI QKVLQQPFYT TDLLFKLVKE SEAMLDQIFP ANETESEIIQ 420
421 AELSEHKFME SLHMKSTIAA LRVLKEIRSG SSTVSVFSLP PLQLNGLDET WKKIPLLEQE 480
481 AK
```

Figure 5.3 Sequence of GST-AtSPX1 and thrombin cleavage site.

A thrombin recognition site is present between GST and AtSPX1 to allow removal of the GST tag. Sequence showing the GST-AtSPX1 fusion protein (56.4 kDa) with the thrombin cleavage site (LVPR//GS) highlighted in yellow. After thrombin cleavage, AtSPX1 (underlined) carries two extra amino acids (GS) at the N terminus (sequence underlined) and has a theoretical Mw of 30.2 kDa.
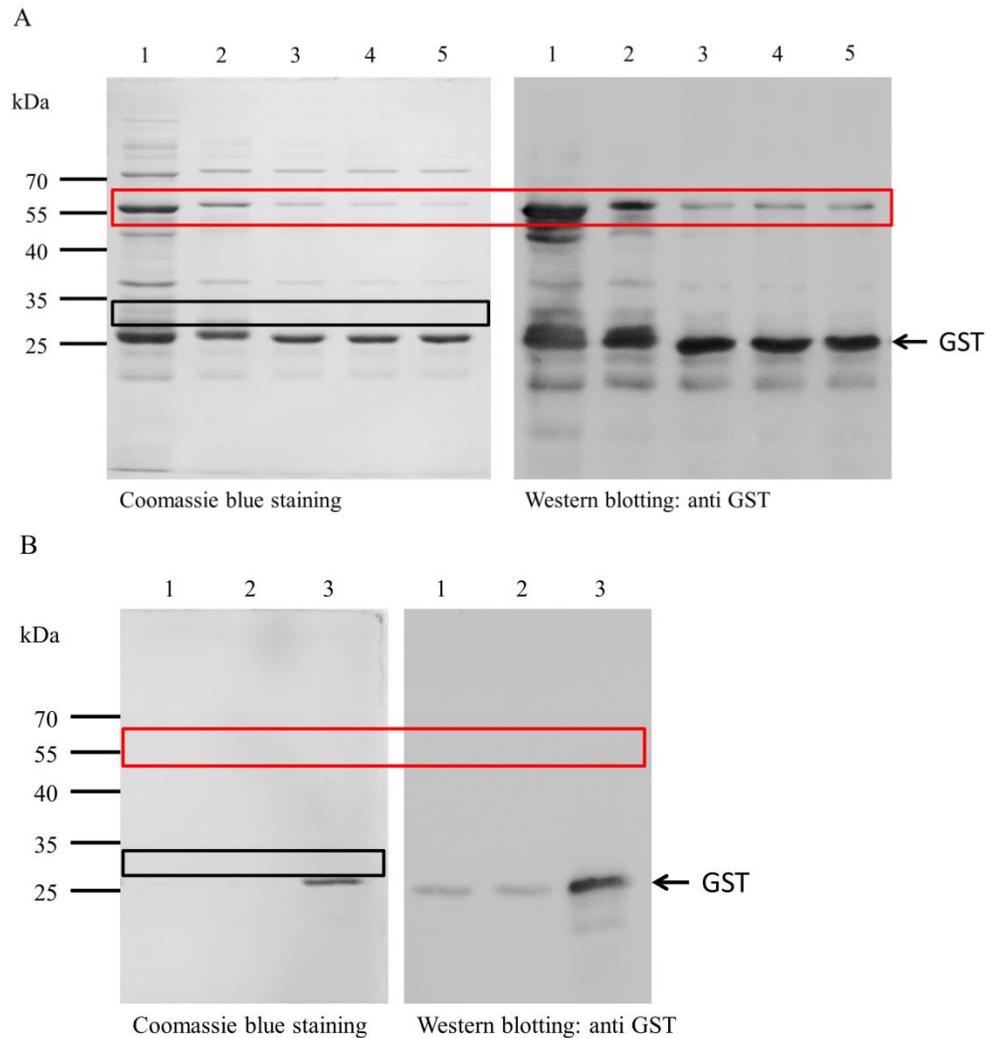
Figure 5.4 SDS-PAGE analysis of GST-AtSPX1 thrombin cleavage and GST tag removal.

Coomassie blue stained gel (left panels) and western blotting (right panels) showing the GST-AtSPX1 thrombin cleavage (A) and GST tag removal using glutathione sepharose resin (B). Affinity purified GST-AtSPX (A1) was buffer exchange into thrombin cleavage buffer (A2) and incubated with thrombin protease at room temperature. Samples after 1 h, 2 h and 4 h of thrombin cleavage (A3-5, respectively) were taken and an equivalent amount of proteins were separated. After 4 h, the cleavage products were subjected to an additional affinity chromatography using glutathione resin. An equivalent amount of proteins from unbound flow-through fraction (B1), column washing fraction (B2) and elution fraction (B3) were separated on 12% SDS-PAGE gels. Equal amount of proteins were also transferred to a nitrocellulose membrane after SDS-PAGE for immunoblotting using anti-GST antibody. Expected positions of GST-AtSPX1 and AtSPX1 are marked by red and black colored boxes, respectively. Free GST protein is indicated by an arrow.

Buffer condition screen for GST-AtSPX1 using Optim system

As the above attempts of further purification or tag removal on GST-AtSPX1 resulted in significant loss of the recombinant protein, it suggests that the buffer condition used in affinity purification may not be optimal for the AtSPX1 protein construct. In order to find a buffer condition that is able to keep the recombinant protein more stable in solution, alternative buffer conditions (Table 5.1) were tested on the affinity purified GST-AtSPX1, using a light-scattering aggregation assay with an Optim 1000 instrument. Static Light Scattering (SLS) at 266 nm was monitored during temperature increase to indicate the aggregation of proteins (Figure 5.5 A-D).

Table 5.1 Alternative buffer conditions screened for GST-AtSPX1 with Optim 1000

Alternated conditions	Parameters
pH range	3.0 – 9.5 by 0.5 increment
Glycerol concentration	0, 2.5%, 5%, 10%, 15 %, 20 %, 25%, 30%
NaCl concentration	0, 50 mM, 150 mM, 250 mM, 500 mM, 750 mM, 1 M, 1.5 M
Additives	ammonium sulphate, potassium acetate, potassium glutamate, potassium chloride, sodium sulphate, sodium chloride, lithium chloride, magnesium chloride, magnesium nitrate, calcium chloride, lithium nitrate, sodium thiocyanate, lithium perchlorate, arginine
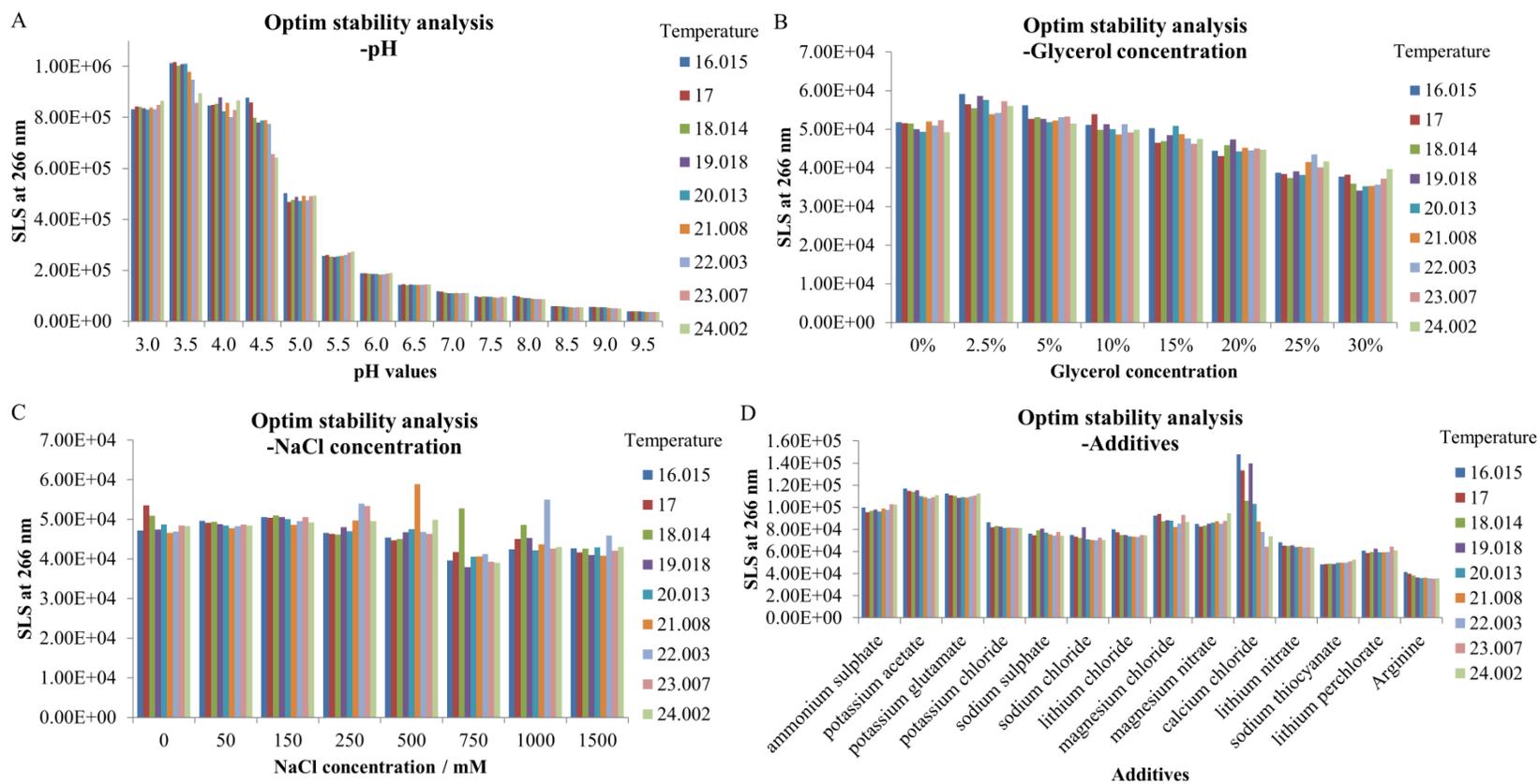
Figure 5.5 Static Light Scattering of GST-AtSPX1 using Optim 1000 buffer conditions screen.

Static Light Scattering (SLS) at 266 nm was monitored during temperature rising to indicate the aggregation of proteins in a series of different buffer conditions, including different pH (A), glycerol concentration (B), NaCl concentration (C) and different additives (D). SLS values at varying temperatures were plotted against a total of 45 different buffer conditions.

As shown in Figure 5.5, GST-AtSPX1 generated less static light scattering in a high pH environment as well as in the presence of a high concentration of glycerol (Figure 5.5 A, B), suggesting the protein is more stable under these conditions. On the contrary, altering NaCl concentration did not affect the protein stability (Figure 5.5 C). The additive screen showed that within the range of screened additives, arginine has a higher possibility of keeping affinity purified GST-AtSPX1 from precipitating (Figure 5.5 D). Although it is of note that due to the high pKa value of arginine, the stabilizing effect from adding arginine could be caused by increased pH.

Although a high pH environment seems to be preferential for GST-AtSPX1 stability, this is not optimal either for further protein purification or for the subsequent functional assays. Therefore, PBS, Tris or HEPES buffer containing 25% glycerol at pH 8.0 was used in sequential affinity chromatography and ion-exchange chromatography on GST-AtSPX1 in order to increase the protein stability (Figure 5.6).
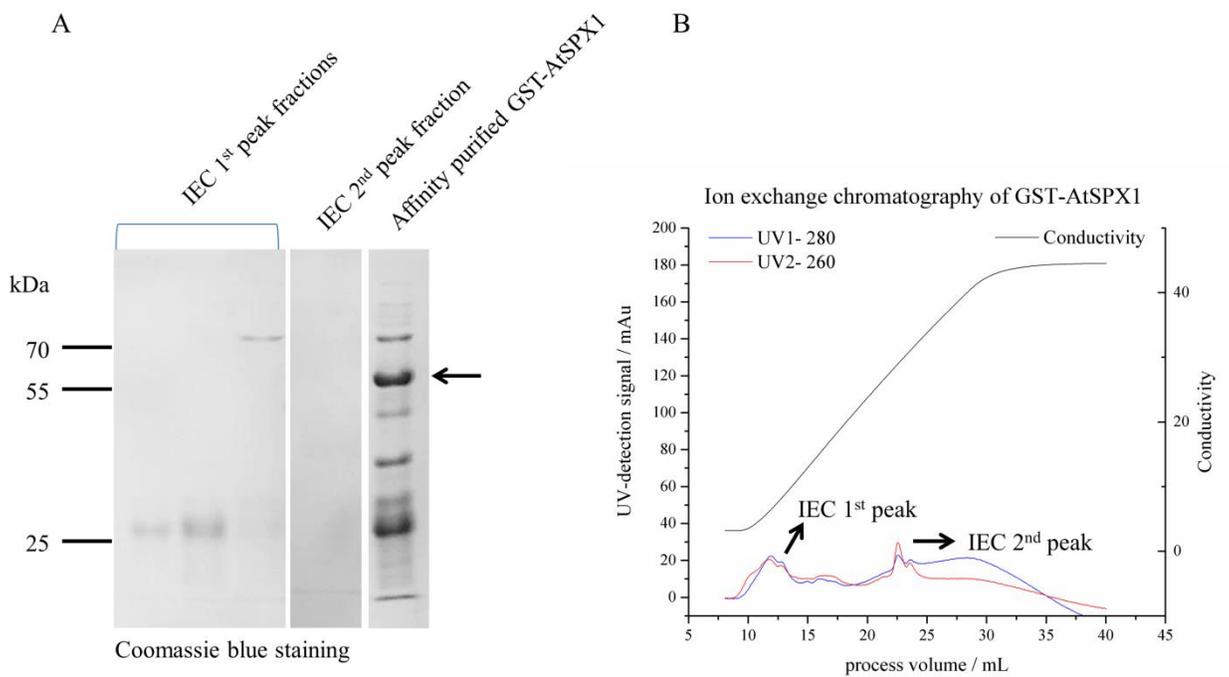
Figure 5.6 GST-AtSPX1 affinity and ion-exchange chromatography in alternative buffer condition (25% glycerol).

GST-AtSPX1 was expressed and purified as described in the previous section. 25% glycerol was included in all the buffers for affinity chromatography as well as for ion-exchange chromatography. A. Equivalent amount of ion-exchange chromatography (IEC) peak fractions were loaded on 12% SDS-PAGE together with the affinity purified GST-AtSPX1 fraction. Proteins were stained with Coomassie blue and GST-AtSPX1 is indicated by an arrow. B. Chromatogram of GST-AtSPX1 IEC with 25% glycerol. Light absorbance at 280 nm and 260 nm are represented by blue and red lines, respectively.

Including 25% glycerol in the buffers did not alter the affinity purification result of GST-AtSPX1, as similar contamination bands are still seen in the purified elution fractions (Figure 5.1, Figure 5.6 A). As the buffer used for ion-exchange purification had a pH that is 1.5 higher than the theoretical pI of GST-AtSPX1 (6.5), the affinity purification enriched GST-AtSPX1 was expected to bind to the Q-HP anion exchange column and subsequently eluted by increased salt concentration. However, the ion-exchange purification step showed no recovery of the GST-AtSPX1 fusion protein in the peak fractions (Figure 5.6 A). Since no absorbance signal at 280 nm was observed in the unbound flow-through and a high column pressure was generated during the ion-exchange chromatography, it is very likely that GST-AtSPX1 was largely precipitating on column during the second purification step. The peak fraction from ion-exchange chromatography is proved to contain only contaminating proteins from the previous purification step, suggesting these proteins rather than GST-AtSPX1 are the reason for low SLS at high glycerol concentration (Figure 5.5 B). Although a second UV absorption peak is also seen in the chromatogram (Figure 5.6 B), no protein was detected on SDS-PAGE (Figure 5.6 A). As this peak has a higher absorbance at 260 nm, it possibly contains some *E. coli* DNA contamination from the affinity chromatography.

As subsequent purification steps or tag removal resulted in significant loss of the recombinant protein and no alternative buffer condition was found to improve the stability or purity of GST-AtSPX1, it suggests the SPX domain-containing protein AtSPX1 is unstable, and tends to aggregate or precipitate especially on its own. Similar poor stability of AtSPX1 and its homologue OsSPX4 has been also reported in recent studies (Wild *et al.*, 2016; Lv *et al.*, 2014). However, when fused to the N-terminal GST tag, the GST-AtSPX1 fusion protein showed a higher solubility (result from previous chapter) and highly consistent purification result from affinity chromatography. As western blotting of the affinity purified GST-AtSPX1 demonstrated most of the faster migrating proteins are either degradation products of GST-AtSPX1 or free GST tag (Figure 5.1), a GST protein was expressed using pGEX-4T-2 plasmid (provided by Dr. Iain Manfield). Affinity chromatography purification of GST was carried out under the same condition as GST-AtSPX1 fusion protein as above. Affinity purified GST protein showed high purity (Figure 5.7) and was therefore used as a negative control in subsequent functional assays.
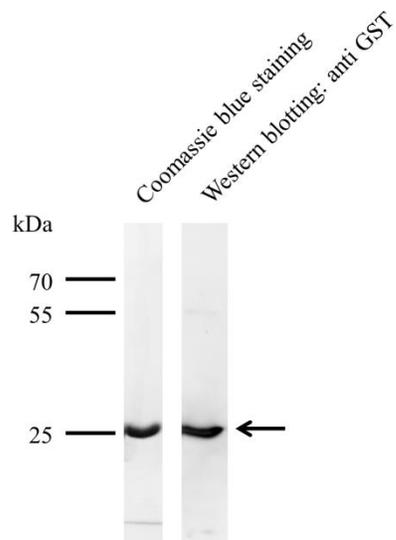
Figure 5.7 SDS-PAGE analysis of affinity chromatography purified GST protein.

Affinity chromatography of GST protein was carried out using glutathione resin on a gravity flow column. Purified protein from elution fraction was separated on 12% SDS-PAGE gel and stained using Coomassie blue (left panel). An equal amount of proteins were also transferred to a nitrocellulose membrane after SDS-PAGE for immunoblotting using anti-GST antibody (right panel). GST (Theoretical Mw 27.4 kDa) is indicated by an arrow.

5.2.2   Characterization of MBP-AtdPHR1

Expression and purification of MBP-AtdPHR1

The transcription factor AtPHR1 is a 'Myb-related' transcription factor which contains a single MYB repeat as well as a predicted coiled-coli domain (Dubos *et al.*, 2010; Rubio *et al.*, 2001). A truncated version of AtPHR1 (AtdPHR1: aa 208-362) containing the DNA-binding domain as well as the coiled-coil domain has been shown to interact with AtSPX1 under Pi sufficient conditions (Puga *et al.*, 2014). To create this truncated version of AtPHR1, the full-length *AtPHR1* coding sequence (At4g28610, ABRC) was used as a template and gene sequence coding amino acid 208-362 was amplified. The PCR product of AtdPHR1 was sub-cloned into pMAL-c2x expression vector to generate an N-terminal MBP fused AtdPHR1 construct (Figure 5.8). The MBP-AtdPHR1 construct used for expression was confirmed by sequencing.

Expression of MBP-AtdPHR1 was carried out in the *E. coli* BL21-gold (DE3) cell strain according to the 'pMAL protein fusion and purification system instruction manual' (New England Biolabs, #E8200S). MBP-AtdPHR1 fusion protein was affinity purified using amylose resin according to the manufacturer's instructions (Chapter 2 section 2.2.5.2) (Figure 5.9).

Figure 5.9 shows that the single step affinity chromatography has successfully removed most of the contaminating proteins from host cells, and the MBP-AtdPHR1 fusion protein was eluted with relatively high purity and higher yield (~ 7 mg purified protein per liter culture). However, as AtPHR1 is a transcription factor and the DNA-binding domain is included in the AtdPHR1 construct, an additional ion-exchange chromatography step was carried out on a HiTrap Q-HP column using an ÄKTA Explorer to remove contaminating *E. coli* DNA (Figure 5.10).

A

```
  1 MKIEEGKLVI WINGDKGYNG LAEVGKKFEK DTGIKVTVEH PDKLEEKFPQ VAATGDGPDI  60
 61 IFWAHDRFGG YAQSGLLAEI TPDKAFQDKL YPFTWDAVRY NGKLIAYPIA VEALSLIYNK 120
121 DLLPNPPKTW EEIPALDKEL KAKGKSALMF NLQEPYFTWP LIAADGGYAF KYENGKYDIK 180
181 DVGVDNAGAK AGLTFLVDLI KNKHMNADTD YSIAEAAFNK GETAMTINGP WAWSNIDTSK 240
241 VNYGVTVLPT FKGQPSKPFV GVLSAGINAA SPNKELAKEF LENYLLTDEG LEAVNKDKPL 300
301 GAVALKSYEE ELAKDPRIAA TMENAQKGEI MPNIPQMSAF WYAVRTAVIN AASGRQTVDE 360
361 ALKDAQTNSS SNNNNNNNNN NLGIEGRISE FGSELRPVST TSSNSNNGTG KARMRWTPEL 420
421 HEAFVEAVNS LGGSERATPK GVLKIMKVEG LTIYHVKSHL QKYRTARYRP EPSETGSPER 480
481 KLTPLEHITS LDLKGGIGIT EALRLQMEVQ KQLHEQLEIQ RNLQLRIEEQ GKYLQMMFEK 540
541 QNSGLTKG
```

B
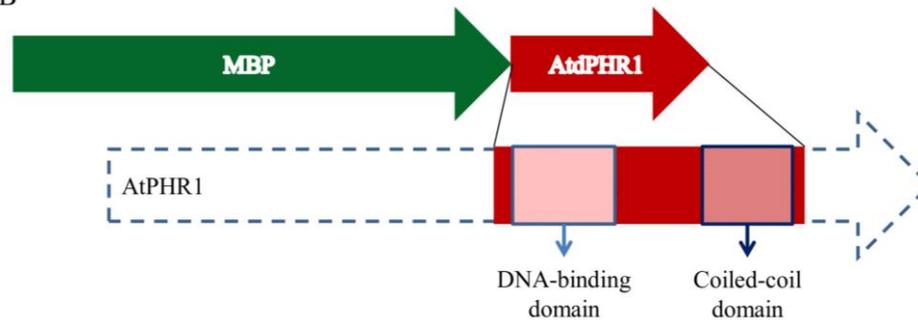
Figure 5.8 Amino acid sequence (A) and sketch diagram (B) for MBP-AtdPHR1 fusion protein.

pMAL-c2x-AtdPHR1 was generated for expression of MBP-AtdPHR1 fusion protein. Gene encoding truncated version of AtdPHR1 (aa 208-362) was sub-cloned into pMAL-c2x expression vector using available *Bam*HI and *Hin*dIII sites. A. Amino acid sequence of MBP-AtdPHR1 fusion protein with MBP and AtdPHR1 sequences highlighted in green and red, respectively. The AtPHR1 DNA-binding domain is underlined and predicted coiled-coil domain is marked with yellow font color. B. Sketch diagram of MBP-AtdPHR1 fusion protein, a myb DNA-binding domain (AtPHR1-aa 227-278) and a predicted coiled-coil domain (AtPHR1-aa 310-355) is included in the AtdPHR1 construct.
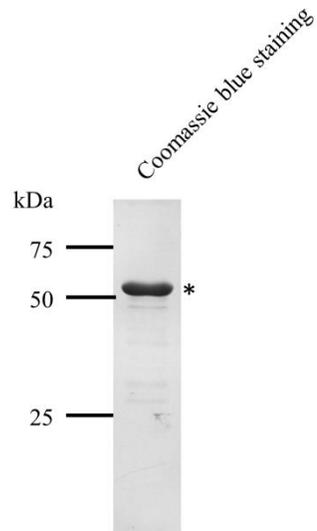
Figure 5.9 SDS-PAGE analysis of affinity chromatography purified MBP-AtdPHR1.

Affinity chromatography of MBP-AtdPHR1 on amylose resin. Purified protein from elution fraction was separated on 12% SDS-PAGE gel and stained using Coomassie blue. MBP-AtdPHR1 (Theoretical Mw 60.7 kDa) is indicated by *.
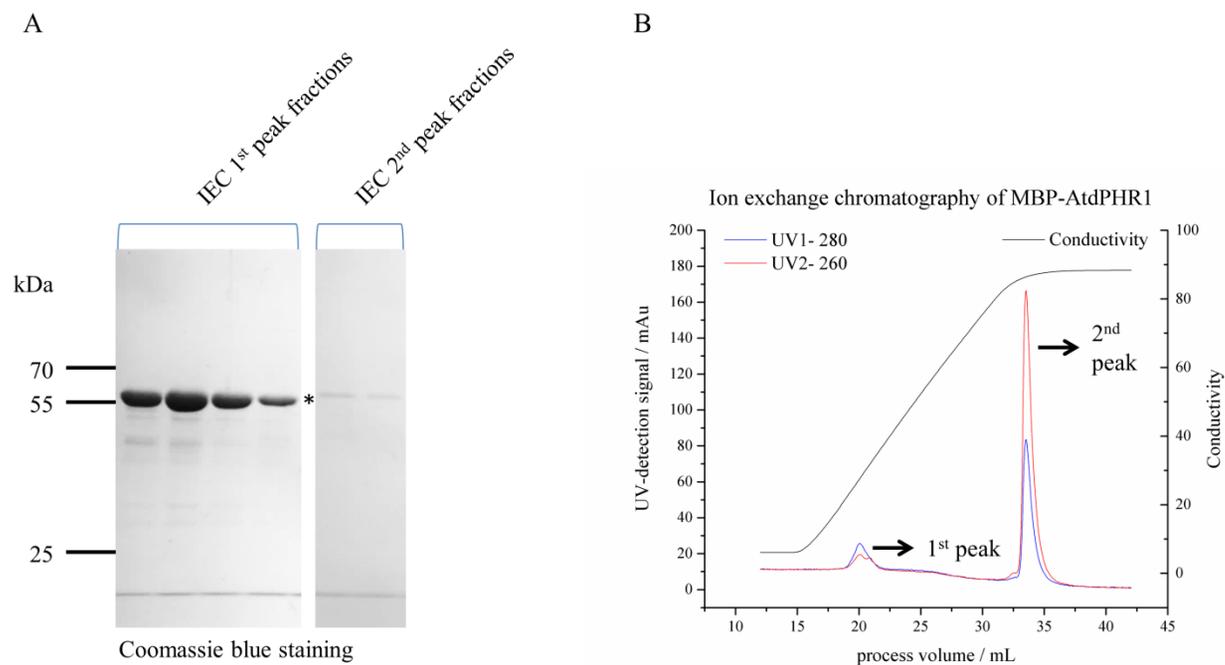
Figure 5.10 MBP-AtdPHR1 ion-exchange chromatography (IEC).

Affinity purified MBP-AtdPHR1 was subjected to a subsequent ion-exchange chromatography step on a HiTrap Q-HP 1 mL column with a gradient from 50 mM to 1 M NaCl in HEPES 10 mM pH 7.5 buffer, using an ÄKTA Explorer. A. SDS-PAGE analysis of MBP-AtdPHR1 IEC fractions. An equivalent amount of protein from each gradient fraction was separated on a 12% SDS PAGE gel and stained using Coomassie blue. MBP-AtdPHR1 (Theoretical Mw 60.7 kDa) is indicated by *. B. Chromatogram of MBP-AtdPHR1 IEC. Light absorbance at 280 nm and 260 nm are represented by blue and red lines, respectively. Arrows indicate the peak positions.

MBP-AtdPHR1 showed high stability during the subsequent ion-exchange chromatography, and the fusion protein was recovered with high purity from the first IEC peak (Figure 5.10 A). Only minor amounts of faster migrating proteins are seen on the gel and are possibly proteolysis products of MBP-AtdPHR1. A second UV-absorption peak towards the end of salt gradient shows higher light absorption at 260 nm than at 280 nm (Figure 5.10 B), indicating the presence of DNA in this peak. This together with the trace amount of MBP-AtdPHR1 protein detected from the 2nd peak (Figure 5.10 A) demonstrates the successful separation of recombinant MBP-AtdPHR1 protein from the *E. coli* host cell DNA.

MBP-AtdPHR1 DNA-binding characterization with Surface Plasmon Resonance (SPR)

The DNA binding activity of recombinant MBP-AtdPHR1 was tested with SPR. DNA probes that contain the specific AtPHR1 recognition site, P1BS motif (Rubio *et al.*, 2001) were taken from the promoter regions of AtSPX1 (2xP1BS motifs) and AtSPX3 (1xP1BS motif), and immobilized on streptavidin (SA) sensor chips to generate approximately 500 RU of immobilized DNA for each flow cell. IEC purified MBP-AtdPHR1 was measured at 2-fold ascending concentrations, ranging from 0.3125 nM to 80 nM for its binding capacity to P1BS motifs (Figure 5.11).

A

```
1xP1BS 5'-ACACTTCGTCACGCTAAAGCTAAGCATATCCGCTTTCATATTCCTTTACACAAC-3'

2xP1BS 5'-CAGAGAAAAAAGGATATTCTAATTAGAAACCTTAAGAATATTCTTTTTAATCCC-3'
```

B
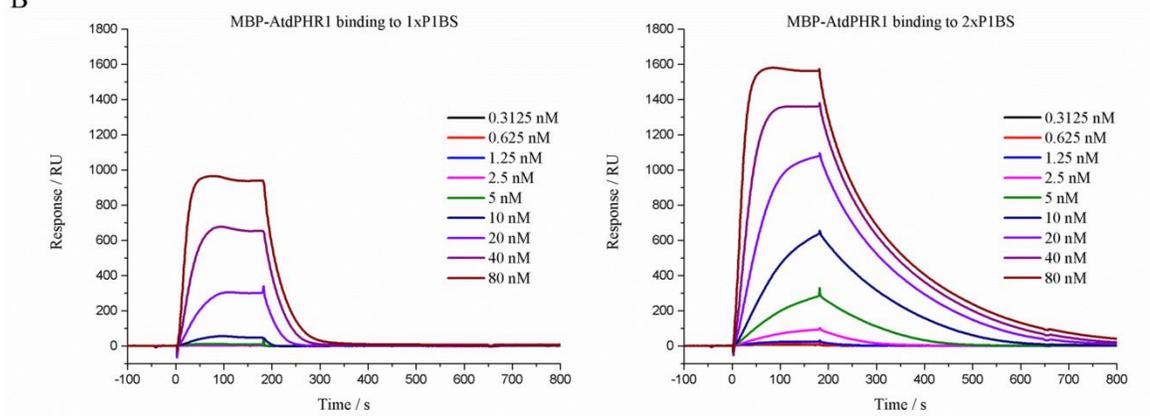
Figure 5.11 MBP-AtdPHR1 DNA-binding characterization using SPR.

DNA probes containing 1xP1BS motif and 2xP1BS motifs (A, P1BS motifs highlighted in yellow) were immobilized on the SPR chip surface and IEC purified MBP-AtdPHR1 was injected onto the chip surface at 2-fold ascending concentrations, ranging from 0.3125 nM to 80 nM. After 3 min injection of MBP-AtdPHR1, SPR buffer was injected to remove bound proteins from the chip surface. The response signal during injection was recorded and plotted against injection time for 1xP1BS (B, left panel) and 2xP1BS (B, right panel) DNA probes.

The binding profiles showed that binding signal increases with increasing MBP-AtdPHR1 concentration, and the MBP-AtdPHR1-DNA interaction reached equilibrium with 40-80 nM MBP-AtdPHR1 (Figure 5.11). The signal at equilibrium at the higher concentrations was approximately twice as high on the 2xP1BS site versus 1xP1BS. As the SPR signal intensity is proportional to the amount of protein bound and as very similar amounts of DNA were immobilized on each flow cell, this result suggests differences in stoichiometry of binding to each DNA probe. Specifically, the 2xP1BS-PHR1 complex appears to have a stoichiometry twice that of the 1xP1BS-PHR1 complex.

In addition, the dissociation rate from the 2xP1BS site was much slower than from the 1xP1BS site (Figure 5.11), showing a higher affinity of MBP-AtdPHR1 to the double binding sites than to the single site. Using a 1:1 model to fit kinetic and affinity parameters to the data, AtdPHR1 binds P1BS sites with high affinity; around K_d of 130 nM for AtdPHR1-1xP1BS and around K_d of 15 nM for the AtdPHR1-2xP1BS complex. This suggests protein cooperativity and dimerization of MBP-AtdPHR1 is occurring during its interaction with P1BS motifs.

Size exclusion chromatography of MBP-AtdPHR1

Dimerization of AtPHR1 was previously observed during DNA association (Rubio *et al.*, 2001). Here we sought to find out if the AtPHR1 dimer forms before or after DNA binding. The protein oligomeric state of MBP-AtdPHR1 in solution was examined at 5 µM concentration using size exclusion chromatography (SEC) (Figure 5.12).
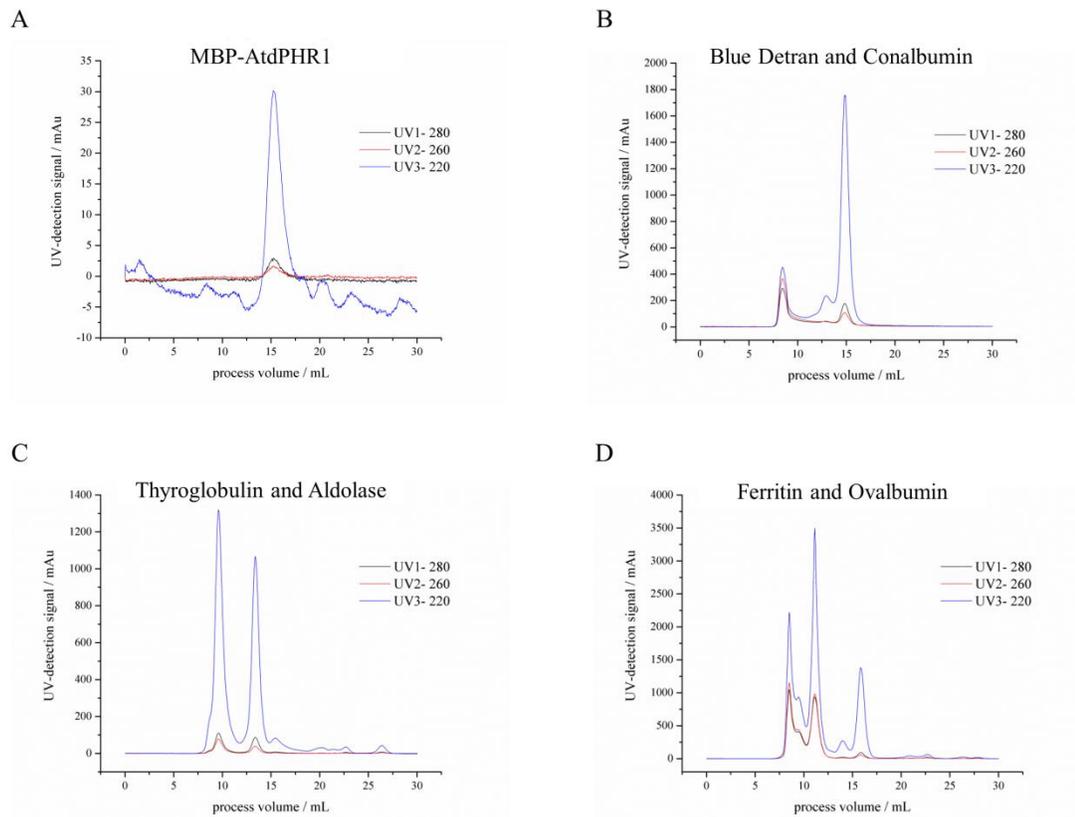
Figure 5.12 Size exclusion chromatography (SEC) of MBP-AtdPHR1 and standard proteins.

5 μM MBP-AtdPHR1 was analyzed on a Superdex 200 10/300 GL column with the same buffer conditions used for SPR. Blue Dextran 2000 and standard proteins with varying molecular weights (Mw) were also analyzed under the same condition. UV-absorption at 220 nm, 260 nm and 280 nm are represented by blue, red and black colored lines, respectively. A. SEC of MBP-AtdPHR1; B. SEC of Blue Dextran 2000 and Conalbumin; C. SEC of Thyroglobulin and Aldolase; D. SEC of Ferritin and Ovalbumin.

A single peak of MBP-AtdPHR1 was seen from SEC (Figure 5.12 A), corresponding to a single oligomeric state of this protein in solution at concentration of 5 μM. As the void volume of the column was determined to be 8.4 mL using Blue Dextran 2000, the partition coefficient Kav for individual standard protein was calculated using equation 5.1 (Table 5.2).

$$Kav = (V_R-V_O)/(V_C-V_O) \hspace{3cm} [5.1]$$

(V_O = void volume, 8.4 mL; V_R = elution volume of proteins; V_C = column geometric bed volume, 24 mL)

A calibration curve was plotted using the linear correlation between Log Mw and Kav values for standard proteins (Figure 5.13) to determine the Mw of MBP-AtdPHR1.

As a linear regression equation was obtained from the standard proteins (R^2>0.95), the molecular mass of MBP-AtdPHR1 peak seen in Figure 5.12 A could be confidently calculated to be 62 kDa. This corresponds to a monomeric state of MBP-AtdPHR1 (Theoretical Mw 60.7 kDa) which shows the transcription factor AtPHR1 is a monomeric protein in solution. Given the concentration of MBP-AtdPHR1 used in SPR was about 1000 times lower than that in SEC, the much slower dissociation rate from the 2xP1BS probe than from the 1xP1BS probe clearly suggests MBP-AtdPHR1 only dimerizes when interacting with DNA molecules.

Table 5.2 Partition coefficient of standard proteins for Superdex 200 10/300 GL column.

Standard protein	Molecular weight (Mw) / kDa	Elution volume / mL	Partition coefficient (Kav)
Ovalbumin	44	15.9	0.48
Conalbumin	75	14.9	0.42
Aldolase	158	13.4	0.32
Ferritin	440	11.1	0.17
Thyroglobulin	669	9.6	0.08

Protein Mw calibration curve
(Superdex 200-10/300 GL)

$y = -0.3344x + 1.0418$
$R^2 = 0.9914$

Kav
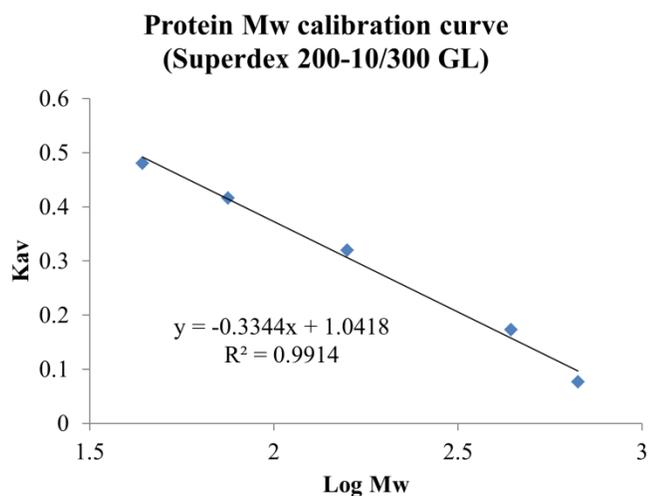
Log Mw

Figure 5.13 Calibration curve for protein Molecular weight (Mw) with Superdex 200 10/300 GL column.

A calibration curve of protein molecular weight was generated by plotting Log Mw against Kav values for standard proteins.

5.2.3   Analysis on Pi-dependent AtSPX1-AtPHR1 interaction

In vitro MBP-AtdPHR1 pull-down assay

It has been demonstrated that GST-AtSPX1 can interact with the truncated transcription factor AtdPHR1 (aa 208-362) when sufficient Pi is present (Puga *et al.*, 2014). However, as protein degradation was observed for GST-AtSPX1, it is important to confirm the interaction is specific between AtSPX1 and AtdPHR1 instead of by contaminating species, in particular the GST tag. Using the above GST-AtSPX1 and MBP-AtdPHR1 constructs, the ability of AtSPX1 to interact with AtdPHR1 was examined with a free GST as a negative control in a qualitative pull down assay (Figure 5.14).

As amylose resin was shown to be efficient at removing endogenous *E. coli* proteins from the MBP-AtdPHR1 over expression strain as well as binding MBP-AtdPHR1 fusion protein, recombinant MBP-AtdPHR1 was immobilized to the amylose affinity matrix and used as a bait protein. The same amount of affinity purified GST-AtSPX1 and GST were incubated with MBP-AtdPHR1 for 2 h at 4°C either with or without 15 mM Pi (Puga *et al.*, 2014), to allow formation of a protein complex. Unbound proteins were then removed by extensive washes and interacting proteins were co-eluted with MBP-AtdPHR1 by maltose containing buffer. The presence of GST-containing proteins in each pull-down fraction was detected by western blotting using anti-GST antibody (Figure 5.15).
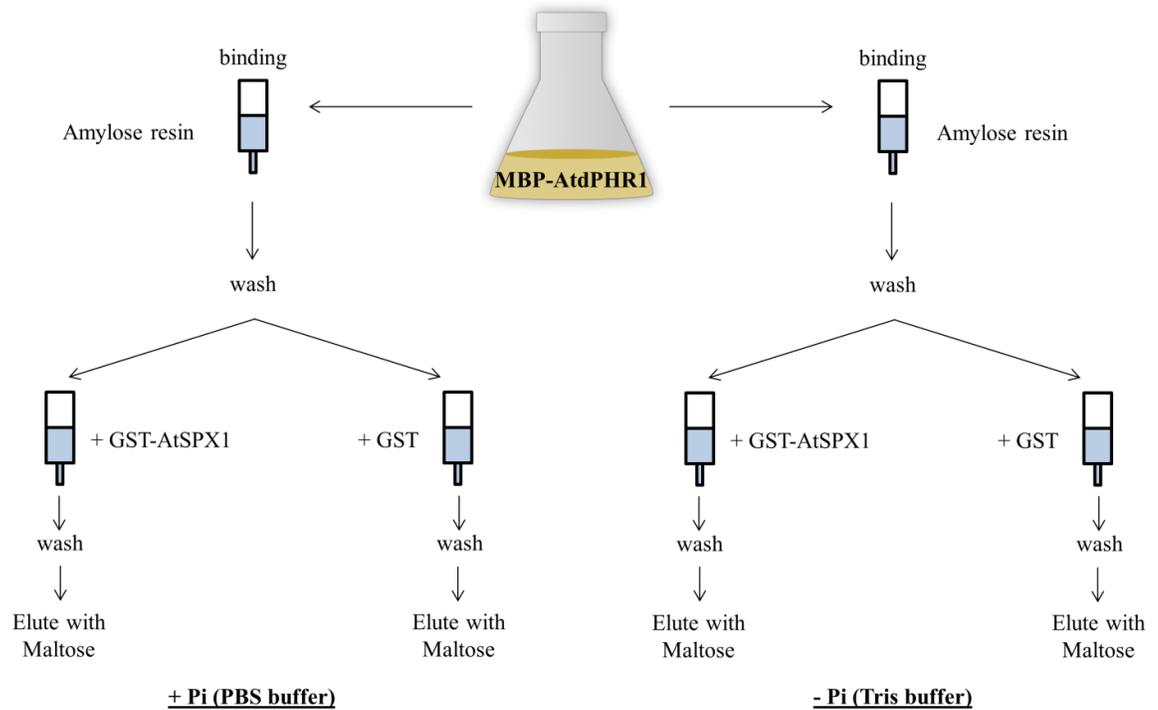
Figure 5.14 Flow chart of *in vitro* **MBP-AtdPHR1 pull-down assay.**

MBP-AtdPHR1 was expressed following the protocol described in materials and methods. After cell lysis, recombinant MBP-AtdPHR1 was bound to amylose affinity matrix as a bait protein and other endogenous proteins from the *E. coli* expression strain were removed by extensive washes. Equivalent amounts of affinity purified GST-AtSPX1 protein was incubated with the immobilized MBP-AtdPHR1 either with or without 15 mM Pi, with GST protein being incubated in the same way as the negative control. Unbound proteins were removed by extensive washing with the binding buffer before interacting proteins were co-eluted with MBP-AtdPHR1 by application of maltose containing buffer.
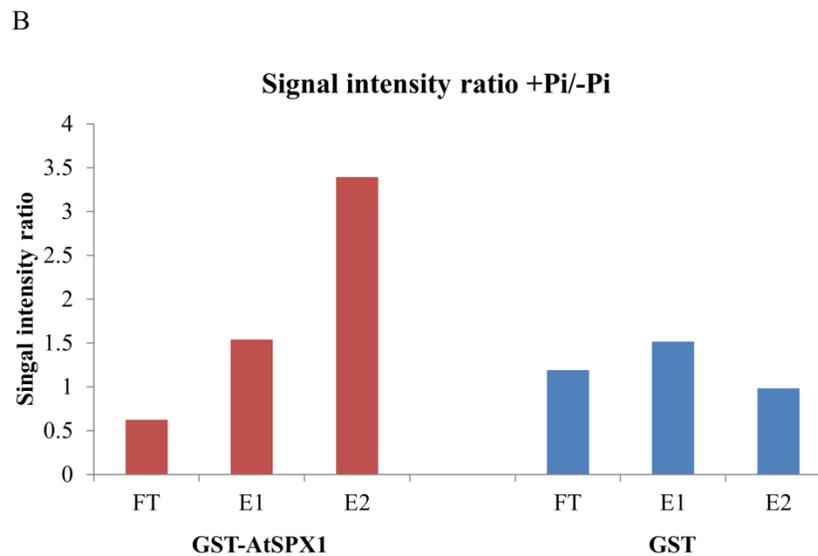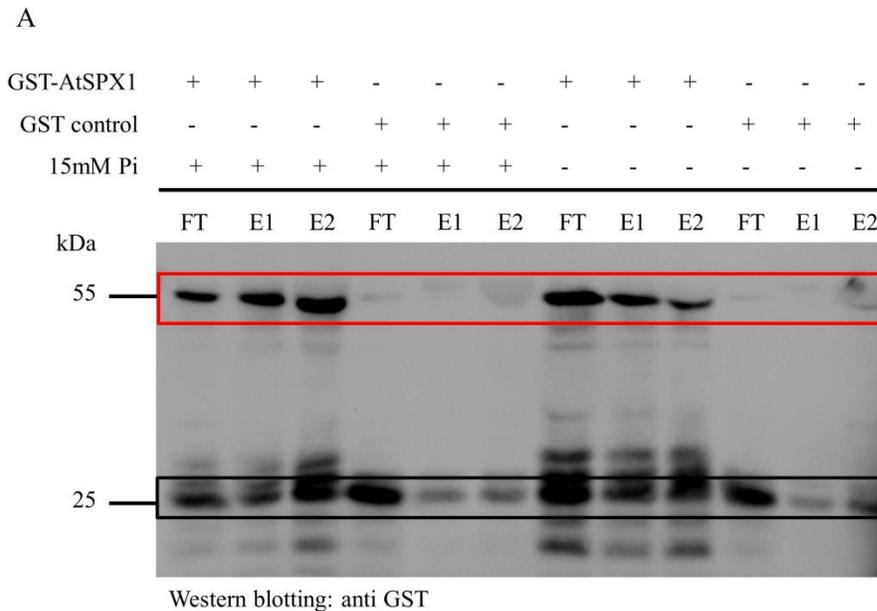
Figure 5.15 Analysis of *in vitro* MBP-AtdPHR1 pull-down assay fractions.

A. Equivalent amount of protein from unbound ('flow through' FT) and the elution fractions (E1 and 2) were separated on 12% SDS-PAGE gel and proteins were transferred to a nitrocellulose membrane after SDS-PAGE for immunoblotting using anti-GST antibody. Expected positions of GST-AtSPX1 fusion protein and free GST protein are marked by red and black colored boxes, respectively; B. Signal intensity of western blotting bands of GST and GST-AtSPX1 was measured in each lane using GeneTool software. The value for each +Pi fraction divided by the corresponding -Pi value was plotted for all pull down fraction.

Western blotting with the anti-GST antibody showed that GST-AtSPX1 interacted with MBP-AtdPHR1 and was co-eluted in the pull-down assay, whereas the negative GST control does not bind to MBP-AtdPHR1 regardless of the Pi status (Figure 5.15 A), suggesting the binding to MBP-AtdPHR1 is a property of the AtSPX1 protein. Western blotting signal intensity comparison also showed that far more GST-AtSPX1 fusion protein is present in the unbound (flow through FT) fraction in the absence of Pi than in its presence (Figure 5.15 B). Conversely 1.5 and 3.5 fold more GST-AtSPX1 protein is present in the elution fractions E1 and E2 in the presence of Pi than in its absence (Figure 5.15 B), indicating Pi has a positive effect on the association of these two proteins. In contrast the ratio of GST in +P and –P fractions is close to 1, indicating no effect of P.

AtSPX1 Pi binding assay

Since the interaction between AtSPX1 and AtdPHR1 is Pi-dependent, I sought to find out whether AtSPX1 directly binds Pi. Both GST and GST-AtSPX1 protein were labeled on the thiol groups using the fluorophore Alexa Fluor 488. Labeled proteins were mixed with unlabeled Pi at 2-fold descending Pi concentrations, ranging from 100 mM to 3 μM. The Pi binding capacity of labeled proteins was determined using Micro Scale Thermophoresis (MST) (Figure 5.16, Figure 5.17). During the MST experiments, the fluorescence of the heat spot in the sample capillaries was monitored to record the temperature gradient triggered thermophoresis of labeled GST-AtSPX1 and GST (Figure 5.16 and Figure 5.17 upper panels). The protein-Pi affinity was quantified by analyzing the fluorescence change after thermophoresis with different concentration of titrated Pi (Figure 5.16 and Figure 5.17 lower panels).

MST data showed both GST and GST-AtSPX1 fusion proteins display Pi-binding activity (Figure 5.16, Figure 5.17). Similar dissociation constants (K_d) for Pi-binding were obtained from GST and GST-AtSPX1 using varying MST laser powers (Table 5.3).

Table 5.3 GST and GST-AtSPX1 Pi-interaction dissociation constant (K_d).

Dissociation constant (K_d) for Pi-binding were obtained for GST and GST-AtSPX1, using different MST laser powers. K_d values for each laser power are expressed as average ($\pm$ SE).

Protein	K_d for Pi (mM)			Average K_d (mM)
	20% laser power	**40% laser power**	**80% laser power**	
GST	11.5 ($\pm$0.54)	13.1 ($\pm$0.73)	10.0 ($\pm$0.79)	11.5 ($\pm$1.55)
GST-AtSPX1	7.86 ($\pm$0.77)	14.9 ($\pm$0.72)	10.8 ($\pm$0.63)	11.2 ($\pm$3.54)
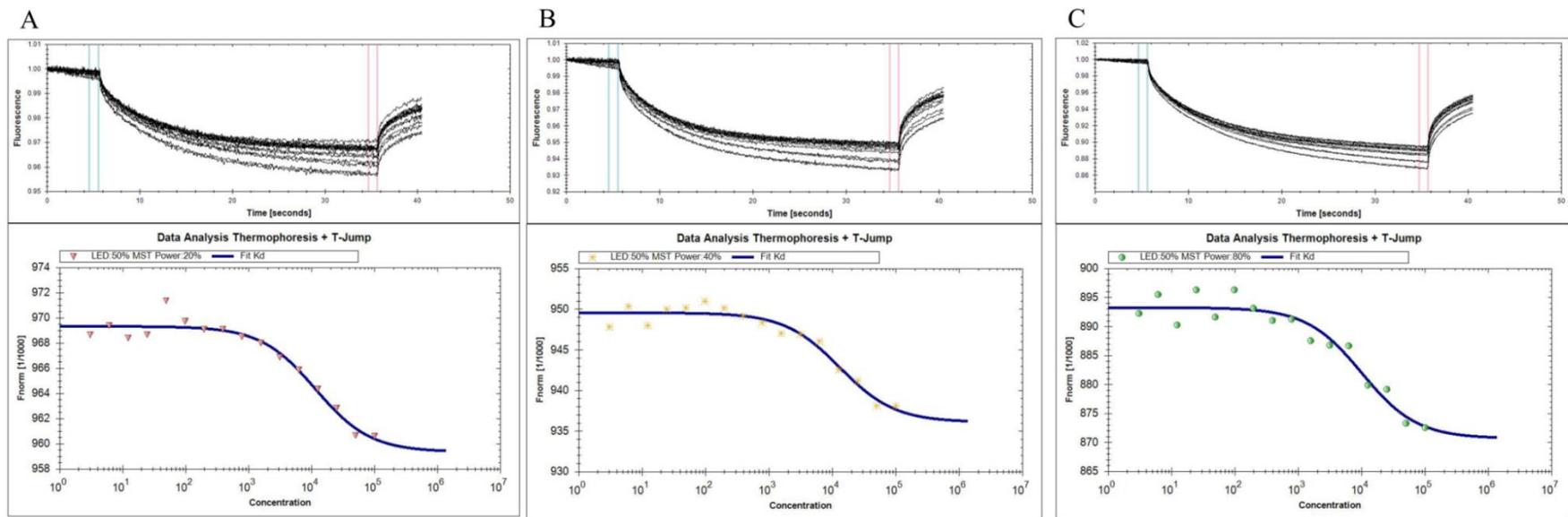
Figure 5.16 GST Pi binding assay with Micro Scale Thermophoresis (MST).

GST was labeled on the thiol groups using fluorophore Alexa Fluor 488. Labelled GST was diluted to give signal intensity ~1500 using 50% LED power. Unlabeled Pi was used to give 2-fold descending final Pi concentrations, ranging from 100 mM to 3 µM for the binding assays. Fluorescence of labeled protein was recorded over time and thermophoresis was plotted against Pi concentration using 20% (A), 40% (B) and 80% (C) MST laser power.
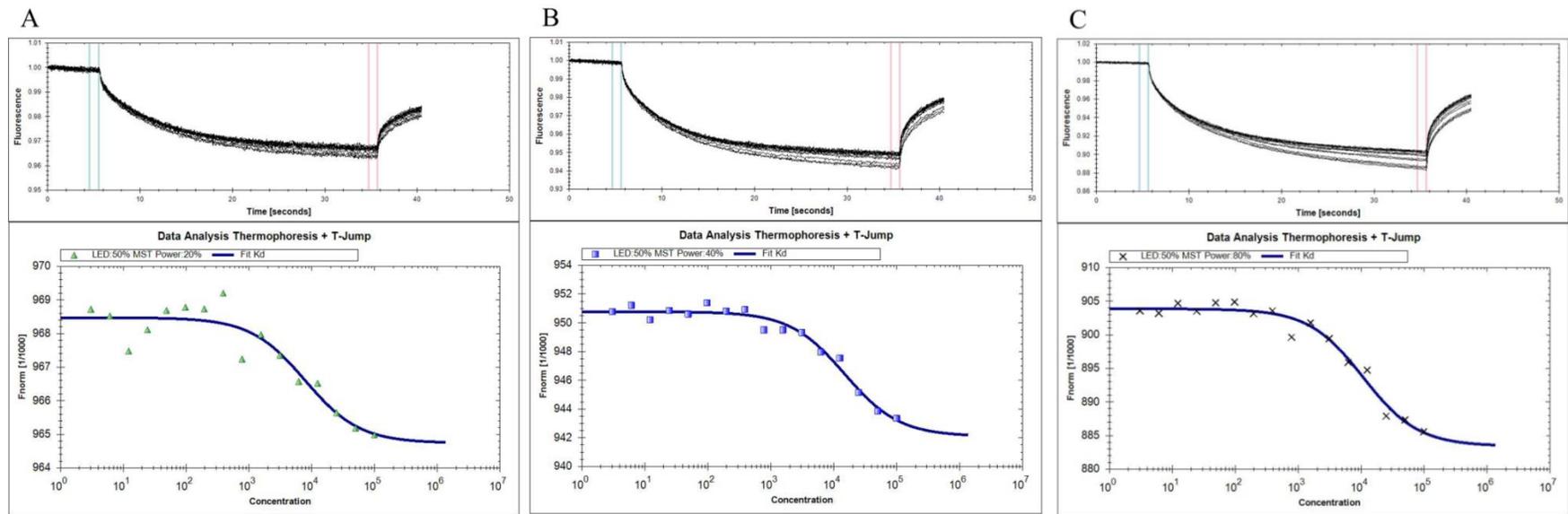
Figure 5.17 GST-AtSPX1 Pi binding assay with Micro Scale Thermophoresis (MST).

GST-AtSPX1 was labeled on the thioyl groups using fluorophore Alexa Fluor 488. Labelled GST-AtSPX1 was diluted to give signal intensity ~1500 using 50% LED power. Unlabeled Pi was used to give 2-fold descending final Pi concentrations, ranging from 100 mM to 3 μM for the binding assays. Fluorescence of labeled protein was recorded overtime and thermophoresis was plotted against Pi concentration using 20% (A), 40% (B) and 80% (C) MST laser power.

As lower laser power is recommended for obtaining good signal to noise ratios, the lowest 20% MST laser power was first tested. A K_d for Pi binding around 8 mM and 11.5 mM was obtained for GST-AtSPX1 and GST control, respectively (Figure 5.16 A, 5.17 A). However, as the difference between bound and unbound state amplitude was lower than 5 response units for GST-AtSPX1, the data suggests higher laser power is required to generate reliable MST signals (MST Starting Guide, NanoTemper technologies). When using 40% and 80% laser power on GST-AtSPX1, ~8 and 20 response units were obtained, respectively. This corresponding to K_d values for Pi at 15 - 11 mM (Figure 5.17 B and C). However, due to the similar K_d detected for GST protein (10-13 mM) (Figure 5.16), the precise affinity of AtSPX1 for Pi could not be determined. It is of note that degradation of AtSPX1 was observed during MST labelling (Figure 5.18). This may lead to the affinity of AtSPX1 binding to Pi being underestimated as the MST florescent label is only on GST protein and any free AtSPX1 that is bound to Pi would not be detected during the experiments. Moreover, as GST protein is shown to be the most dominant contamination in the labeled GST-AtSPX1 used for MST (Figure 5.18), the Pi-binding signal from AtSPX1 could be masked by that from GST. Nevertheless, a very recent study on SPX domains from ScVtc2 and HsXPR1 showed this protein domain, when not fused to a GST tag, can physically bind Pi molecules through its phosphate binding cluster (PBC) (Wild *et al.*, 2016). A similar K_d for Pi between 5-20 mM was also obtained from ScVtc2 and HsXPR1 SPX domains using NMR titrations (Wild *et al.*, 2016), suggesting AtSPX1 is likely to bind Pi in the millimolar range.

Although both GST and GST-AtSPX1 displayed Pi binding and showed very similar K_d for Pi, only GST-AtSPX1 was seen to interact with AtdPHR1 in the presence of Pi. In order to understand how this AtSPX1-AtdPHR1 interaction affects the AtdPHR1 – DNA-binding activity, the interaction mechanism between AtSPX1 and AtdPHR1 was studied using SPR.
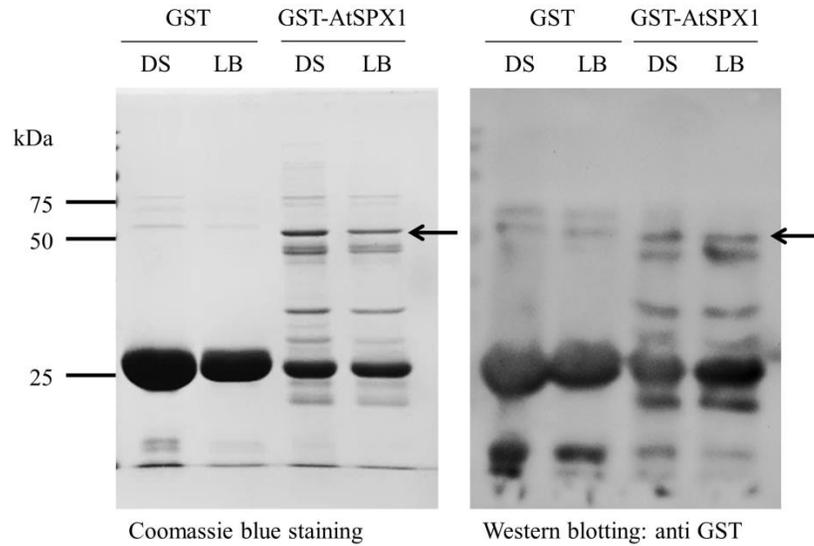
Figure 5.18 SDS-PAGE analysis of GST and GST-AtSPX1 proteins used for MST.

Affinity purified GST and GST-AtSPX1 were dialyzed to remove DTT and Glutathione and labeled on the thiol groups using fluorophore Alexa Fluor 488 for MST. Proteins after dialysis (DS) and labeling (LB) were separated on 12% SDS-PAGE gels and stained using Coomassie blue (left panel). Equal amount of proteins were also transferred to a nitrocellulose membrane after SDS-PAGE for immunoblotting using anti-GST antibody (right panel). Arrows indicate GST-AtSPX1 fusion protein (Theoretical Mw 56.4 kDa).

5.2.4 Investigation of AtSPX1-AtdPHR1 interaction mechanism

Having shown the Pi-dependent interaction between recombinant GST-AtSPX1 and MBP-AtdPHR1 proteins, two different models of interaction between these proteins were examined. In the first binding model, the ability of GST-AtSPX1 to displace MBP-AtdPHR1 from its associating DNA molecules, as proposed by Puga *et al.*, 2014, was tested. In the second binding model, we tested if the formation of a GST-AtSPX1 – MBP-AtdPHR1 protein complex could prevent MBP-AtdPHR1 from binding to DNA molecules.

Testing GST-AtSPX1 – MBP-AtdPHR1 interaction model 1: The ability of GST-AtSPX1 to displace MBP-AtdPHR1 from DNA

The first GST-AtSPX1 – MBP-AtdPHR1 interaction model was tested by sequential injection of these two proteins to the immobilized DNA probes (Chapter 2 section 2.2.6.2). Since the binding equilibrium of both the 1xP1BS and 2xP1BS probes was reached when 80 nM MBP-AtdPHR1 was used in previous SPR assays, the same MBP-AtdPHR1 concentration was used in the sequential injection experiments to make sure binding equilibrium was reached before the MBP-AtdPHR1 – DNA interaction was interrupted. During the SPR experiments, 100 μL MBP-AtdPHR1 at 80 nM was first injected to the SA chip where the DNA probes were immobilized. This was immediately followed by a second injection containing 120 μL either buffer or GST-AtSPX1 at varying concentrations. The SPR result showed that GST-AtSPX1 at different concentrations does not alter either the equilibrium binding signal or the dissociation rate of MBP-AtdPHR1 from the DNA (Figure 5.19). This is true for both the 1x and 2xP1BS binding site probes regardless of the Pi availability (Figure 5.19). Different concentrations of GST-AtSPX1 showed similar effect to buffer alone, indicating that GST-AtSPX1 cannot remove MBP-AtdPHR1 from the associated P1BS motifs. A negative control sequential injection containing SPR buffer and the highest concentration of GST-AtSPX1 only displayed a low signal that is close to base line (Figure 5.19), showing the GST-AtSPX1 protein does not bind to P1BS motifs and the low signal is probably caused by non-specific protein binding to the chip surface.
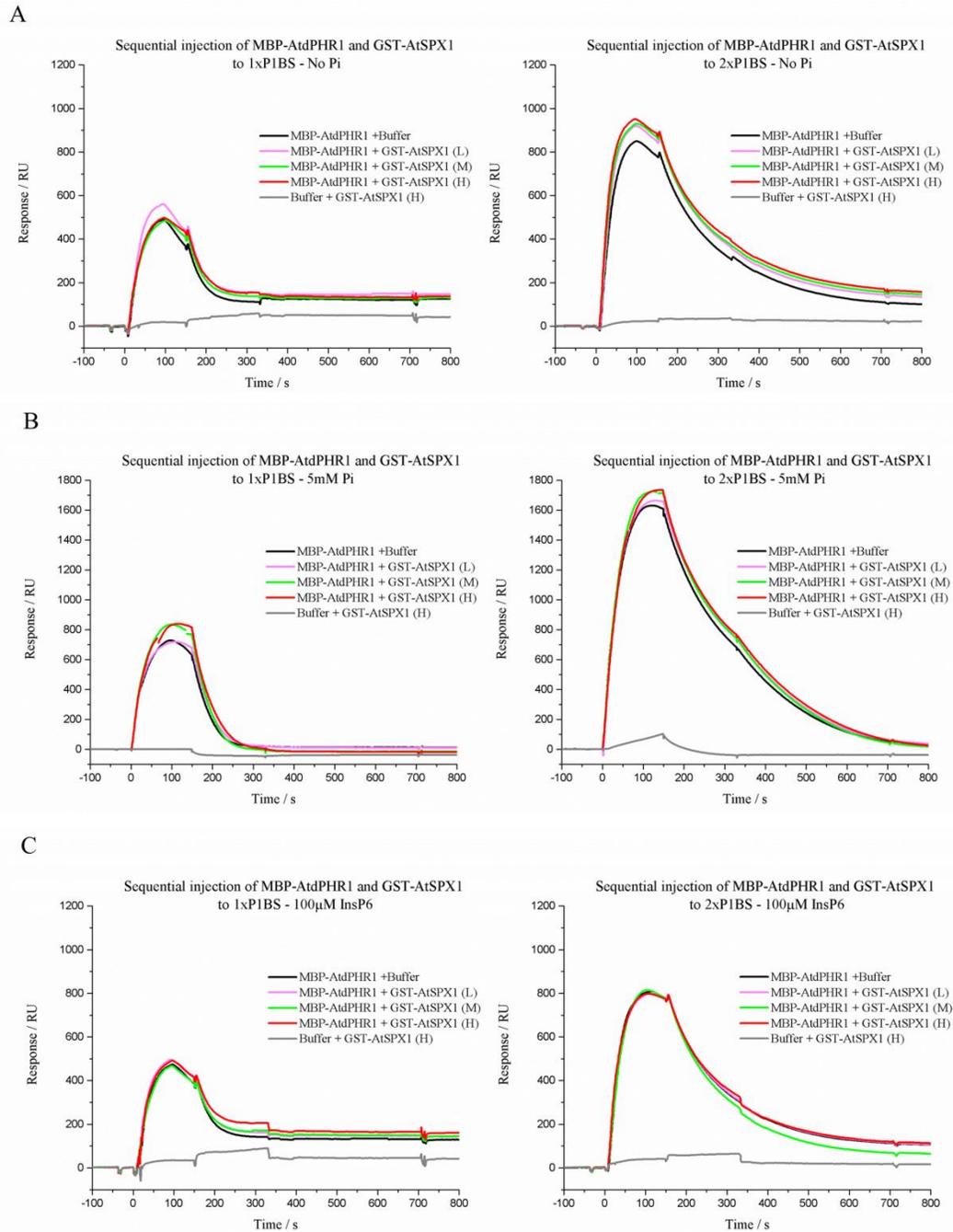
Figure 5.19 Sequential injection of MBP-AtdPHR1 and GST-AtSPX1 to DNA probes.

SPR plots of sequential injection of MBP-AtdPHR1 and GST-AtSPX1 to immobilized DNA probes containing 1xP1BS (left panels) and 2xP1BS (right panels) with no Pi (A), 5 mM Pi (B) or 100 μM InsP6 (C). Each SPR experiment contained 5 injection cycles, commencing 80 nM MBP-AtdPHR1 + Buffer, followed by 80 nM MBP-AtdPHR1 + GST-AtSPX1 (Low, ~40 nM), 80 nM MBP-AtdPHR1 + GST-AtSPX1 (Medium, ~200 nM), 80 nM MBP-AtdPHR1 + GST-AtSPX1 (High, ~1000 nM) and Buffer + GST-AtSPX1 (High, ~1000 nM).

GST-AtSPX1 – MBP-AtdPHR1 interaction model 2: A pre-formed GST-AtSPX1 – MBP-AtdPHR1 complex prevents AtPHR1 from binding to DNA

The second GST-AtSPX1-MBP-AtdPHR1 interaction model was tested by pre-mixing the two proteins prior to the injection to DNA probes (Chapter 2 section 2.2.6.2). 80 nM MBP-AtdPHR1 was pre-mixed with either SPR buffer or varying concentrations of GST-AtSPX1 and incubated on ice for 30 min to allow protein complexes to form before each experiment. 120 μL protein mix was then injected to the SA chip where the DNA probes were immobilized. The SPR results showed a higher equilibrium signal and slower dissociation rate for double P1BS sites compared to single site probes (Figure 5.20), as seen in the previous MBP-AtdPHR1 DNA-binding SPR results (Figure 5.11).

Without Pi, a slightly higher binding signal was seen when MBP-AtdPHR1 was pre-mixed with low (~40 nM) and medium concentration (~200 nM) of GST-AtSPX1. This is probably caused by non-specific protein binding which is also seen in the negative control of pre-mixed SPR buffer and GST-AtSPX1 (Figure 5.20 A). However, the equilibrium binding signal of MBP-AtdPHR1 was decreased by pre-mixing MBP-AtdPHR1 with a high concentration (~1000 nM) of GST-AtSPX1 in the presence of 5 mM Pi (Figure 5.20 B). No significant change of dissociation rate of MBP-AtdPHR1 and P1BS motifs was observed during the pre-mixing binding, suggesting that GST-AtSPX1 can only interact with free MBP-AtdPHR1 and that GST-AtSPX1 and P1BS motifs may share the same or very close binding sites on MBP-AtdPHR1. A higher proportion of binding signal loss is also seen from the single P1BS probes (Figure 5.20 B), indicating multiple P1BS motifs could stabilize the interaction between MBP-AtdPHR1 and DNA molecules.

During the course of this work the SPX domains from ScVtc2 and HsXPR1 were reported to bind InsP6 with a higher affinity than Pi. Therefore we tested the effect of InsP6 on the GST-SPX1 – MBP-AtdPHR1 interaction using both models. When the GST-AtSPX1 and MBP-AtdPHR1 were pre-mixed in the presence of 100 μM InsP6, GST-AtSPX1 did not alter the MBP-AtdPHR1 – DNA dissociation rate or reduce the equilibrium MBP-AtdPHR1 binding signal (Figure 5.19 C, 5.20 C). However, when the concentration of InsP6 was increased to 500 μM, pre-mixing with a high concentration (~1000 nM) of GST-AtSPX1 was able to decrease the equilibrium binding signal of MBP-AtdPHR1, with a greater signal reduction from the single P1BS motif (Figure 5.20 D), similar to the effect

caused by 5 mM Pi (Figure 5.20 B). It is of note that mass spectrometry analysis showed the InsP6 reagent used for this experiment contains other inositol phosphate components (InsP5, InsP4 and InsP3) (Figure 5.21) that do not have affinity for the SPX domain as high as that of InsP6 (Wild *et al.*, 2016). Therefore the concentration of active InsP6 in this experiment is overestimated.

A

Pre-mixture of MBP-AtdPHR1 and GST-AtSPX1
injection to 1xP1BS - No Pi

Pre-mixture of MBP-AtdPHR1 and GST-AtSPX1
injection to 2xP1BS - No Pi

MBP-AtdPHR1 +Buffer
MBP-AtdPHR1 + GST-AtSPX1 (L)
MBP-AtdPHR1 + GST-AtSPX1 (M)
MBP-AtdPHR1 + GST-AtSPX1 (H)
Buffer + GST-AtSPX1 (H)

B

Pre-mixture of MBP-AtdPHR1 and GST-AtSPX1
injection to 1xP1BS - 5 mM Pi

Pre-mixture of MBP-AtdPHR1 and GST-AtSPX1
injection to 2xP1BS - 5 mM Pi

MBP-AtdPHR1 +Buffer
MBP-AtdPHR1 + GST-AtSPX1 (L)
MBP-AtdPHR1 + GST-AtSPX1 (M)
MBP-AtdPHR1 + GST-AtSPX1 (H)
Buffer + GST-AtSPX1 (H)

C

Pre-mixture of MBP-AtdPHR1 and GST-AtSPX1
injection to 1xP1BS - 100μM InsP6

Pre-mixture of MBP-AtdPHR1 and GST-AtSPX1
injection to 2xP1BS - 100μM InsP6

MBP-AtdPHR1 +Buffer
MBP-AtdPHR1 + GST-AtSPX1 (L)
MBP-AtdPHR1 + GST-AtSPX1 (M)
MBP-AtdPHR1 + GST-AtSPX1 (H)
Buffer + GST-AtSPX1 (H)

Figure 5.20 Pre-mixed MBP-AtdPHR1 and GST-AtSPX1 injection to immobilized DNA probes.

SPR plots of premixed MBP-AtdPHR1 and GST-AtSPX1 samples injected to immobilized DNA probes containing 1xP1BS (left panels) and 2xP1BS (right panels) in the presence of no phosphate (A), 5 mM Pi (B), 100 μM InsP6 (C) or 500 μM InsP6 (D). Each SPR experiment contained 5 injection cycles of each of the following mixtures: 80 nM MBP-AtdPHR1 + Buffer, 80 nM MBP-AtdPHR1 + Low concentration (~40 nM) of GST-AtSPX1, 80 nM MBP-AtdPHR1 + Medium concentration (~200 nM) of GST-AtSPX1, 80 nM MBP-AtdPHR1 + High concentration (~1000 nM) of GST-AtSPX1 and Buffer + High concentration (~1000 nM) of GST-AtSPX1. Mixed samples were incubated on ice for 30 min before injected to the immobilized DNA probes.
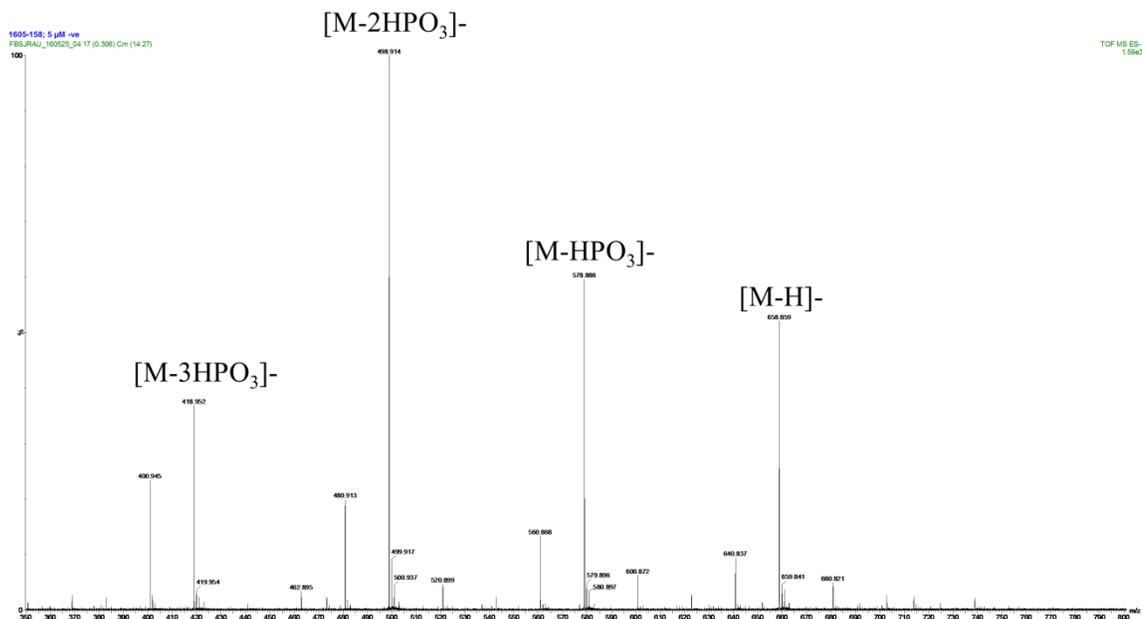
Figure 5.21 Mass spectrometry analysis of inositol phosphate (InsP6) used in this experiment.

Inositol phosphate (InsP6) was analyzed in negative ion nanoelectrospray mode. Signal at 658.859 ([M-H]-) indicates the intact, deprotonated molecule of InsP6. Other peaks observed in the spectrum correspond to loss of the phosphate (HPO_3) groups (InsP5: [M-HPO_3]-; InsP4: [M-2HPO_3]-; InsP3: [M-3HPO_3]-).

5.3 Discussion

When facing low environmental Pi stress, plants adopt multiple strategies. Many transcription factors have been shown to play critical roles in plant Pi regulation pathways (López-Arredondo *et al.*, 2014). In *Arabidopsis*, it has been established that a subset of PSI genes, including the high affinity Pi transporter AtPHT1 family members are regulated by the transcription factor AtPHR1 through interaction with the P1BS motifs in the promoter regions of the target genes (Rubio *et al.*, 2001). Using a truncated version of AtPHR1, which is directly equivalent to the one used by Puga *et al.*, 2014, the SPR data suggests AtPHR1 binds to P1BS motifs as a dimer, which is in consistent with previous study using Electrophoresis Mobility Shift Assay (EMSA) (Rubio *et al.*, 2001). In the same study, Rubio *et al.* also defined the PHR1 protein as a Myb-CC family member for possessing both the Myb-DNA binding domain and a predicted coiled-coil domain (aa 310-355) (Rubio *et al.*, 2001). As the intact coiled-coil domain from AtPHR1 has been shown to be important for the high-affinity P1BS motif binding (Rubio *et al.*, 2001), it is also likely to be involved in the dimerization of AtPHR1. Studies on the rice homologue OsPHR2 have also demonstrated the formation of an OsPHR2 homodimer in the nucleus (Lv *et al.*, 2014), possibly through the coiled-coil domain. However, unlike the OsPHR2, we found that when not associated with DNA, AtPHR1 exists as a monomer in solution, despite the construct used having the full coiled-coil domain and being shown to dimerize on DNA. Structural study on AAR10-B, the DNA binding domain of another *Arabidopsis* Myb-related transcription factor, showed this domain interacts with its DNA target by binding to the first four bases (Hosoda *et al.*, 2002). A model based on this study suggests that AtPHR1 could interact with DNA as a dimer by binding two recognition helices of AtPHR1 to the major groove of the P1BS motif in opposite directions, each recognizing half of the sequence (Hosoda *et al.*, 2002). Therefore, based on this model, the dissociation from target DNA would cause the dissociation of AtPHR1 dimers. The SPR results also show the dimerization of AtPHR1 is dependent on the available P1BS motifs in the DNA molecule, with the interaction being stabilized by multiple P1BS motifs from the same DNA molecule. As similar dimerization generated protein cooperativity has been shown to be of great importance for *in vivo* DNA-binding functions (Roeland Boer *et al.*, 2014), the arrangement of P1BS motifs in the promoter region might also be used for DNA expression pattern prediction under Pi limited conditions.

Since AtPHR1 itself is not transcriptionally regulated in response to phosphate level and preserves a nuclear localization independently of Pi availability (Rubio *et al.*, 2001), an inhibitor of AtPHR1 or a transcription corepressor is needed to turn off the expression of PSI genes under Pi sufficient conditions. A recent study in *Arabidopsis* has shown that the SPX exclusive proteins AtSPX1/AtSPX2 can interact with AtPHR1 protein in the presence of Pi, resulting in reduced *in vitro* AtPHR1 binding to P1BS under ascending Pi concentrations (Puga *et al.*, 2014). Conversely P1BS could only compete with AtSPX1 for AtPHR1 in the absence of Pi (Puga *et al.*, 2014). Similar results were reported for OsPHR2 and OsSPX1, OsSPX2 and OsSPX4 (Wang *et al.*, 2014; Lv *et al.*, 2014), pointing to an inhibition effect of SPX proteins on PHR1 transcription factors. So far, it has only been shown in rice that the OsSPX4 could partially inhibit OsPHR2 by retaining it in the cytosol (Lv *et al.*, 2014). However, the mechanism of PHR1 being inhibited from binding to DNA in the nucleus has not been demonstrated. In *Arabidopsis*, it was stated that 'EMSAs showed that, in the presence of Pi, GST-SPX effectively displaced the ΔPHR1/P1BS interaction.' (Puga *et al.*, 2014). However the mechanism of inhibition of PHR1 binding to P1BS was not demonstrated, and SPR data from this chapter clearly shows that mechanistically speaking GST-AtSPX1 does not 'displace the ΔPHR1/P1BS interaction'. The results show that AtSPX1 inhibits DNA binding by MBP-AtdPHR1, rather than removing transcription factor that is already bound to DNA.

Although the importance of these SPX domain-containing proteins has been well recognized in plant Pi regulation (Wang *et al.*, 2009a; Wang *et al.*, 2009b; Stefanovic *et al.*, 2011; Duan *et al.*, 2008; Secco *et al.*, 2012a), it is not until very recently that any structural information has become available, possibly due to the poor stability of proteins containing this domain (Wild *et al.*, 2016). A study on a rice homologue demonstrated the quick turnover of OsSPX4 by the proteasome pathway, although the protein stability can be enhanced by high level of Pi or phosphite (Lv *et al.*, 2014). Similar instability and protein degradation has been observed in the expression and purification of AtSPX1 in this project, and despite the many different constructs tested and many alternative conditions screened (Chapter 4 section 4.2.3, Chapter 5 section 5.2.1), AtSPX1 still showed high instability, especially on its own. Nevertheless, since the N-terminal GST fused construct showed higher solubility and is relatively more stable, we have shown that the Pi-dependent interaction with AtPHR1 is specific to AtSPX1, using free GST protein as a negative

control. This indicates the recombinant AtSPX1 protein obtained is correctly folded and biologically active. In spite of the instability of many SPX proteins, a structural study on certain SPX domains has proved to be achievable (Wild *et al.*, 2016). Using the SPX domains from ScVtc2 and HsXPR1, NMR titrations showed a K_d for Pi between 5-20 mM (Wild *et al.*, 2016). Microscale thermophoresis (MST) data on GST-AtSPX1 also showed a similar K_d for Pi around 10 mM. However, despite of the lack of previous evidence of Pi-binding for GST tags, a similar K_d for Pi was detected for GST protein. Due to the presence of GST as a major contamination, the precise affinity of AtSPX1 for Pi could not be determined. Nevertheless, when 5 mM Pi was used in the GST-AtSPX1 – MBP-AtdPHR1 interaction, it was sufficient for AtSPX1 to prevent AtdPHR1-DNA binding, showing the conservation of this protein domain in Pi binding capacity. On the other hand, while a K_d of around 50 μM was reported for InsP6 binding from the OsSPX4/OsPHR2 complex (Wild *et al.*, 2016), 100 μM InsP6 did not promote the interaction between AtSPX1 and AtdPHR1. The AtdPHR1-DNA interaction was only inhibited by AtSPX1 with 500 μM InsP6, suggesting while SPX domains generally show a higher affinity to InsP6 than to Pi, InsP6 binding ability of SPX domain could vary among closely related homologues. However, given the purity of InsP6 used for this experiment was to a certain extent compromised by the loss of phosphate groups, the concentration of active InsP6 is likely to be lower than reported. Therefore the affinity of AtSPX1 to InsP6 might be higher. This suggests InsP6, being the most abundant plant InsP (Desai *et al.*, 2014), is likely to serve as a direct intracellular Pi signalling molecule to AtSPX1. On the other hand, although the cellular Pi concentration is typically around mM range, the concentration of free cytosolic Pi is found to be much lower (Pratt *et al.*, 2009). In spite of the difficulty of accurately measuring the concentration of cytosolic Pi, using ^{31}P-NMR and optimized perfusion parameters, Pratt *et al.* were able to separate the cytosolic Pi signal from the mitochondria and plastid Pi signals, and reported a cytosolic Pi concentration at μM range in *Arabidopsis* cells (Pratt *et al.*, 2009). Given the fact that the observed affinity between SPX domains and Pi is at mM range (Wild *et al.*, 2016) (this study), it suggests that Pi itself may not be directly sensed by SPX domains in cytosol. However, in the nucleus, where the concentration of Pi is much higher than that of the cytosol (Libanati and Tandler, 1969), it is still possible for Pi molecules to conduct signals to SPX domain and/or to promote the interaction between AtSPX1 and AtPHR1.

Since the SPR data shows GST-AtSPX1 cannot disrupt MBP-AtdPHR1 – DNA interaction but was able to form a protein complex with free MBP-AtdPHR1 in the presence of Pi / InsP6 and prevent MBP-AtdPHR1 from interacting with DNA, we demonstrate that AtSPX1 could modulate the transcriptional regulation of PSI genes by tuning the dynamic equilibrium of the AtPHR1-P1BS interaction (Figure 5.22). It is significant that as AtSPX1 is itself a downstream target of AtPHR1, the inhibition effect of AtSPX1 on AtPHR1-DNA interaction could also regulate its own expression and therefore generate a negative feedback loop. Since AtPHR1 was shown to bind to DNA molecules as a dimer, the fact that AtSPX1 cannot remove AtdPHR1 from DNA association indicates the AtSPX1 binding site from AtPHR1 is either occupied or buried in the AtPHR1 dimer – DNA complex. Analysis of the specific DNA-interacting residues in ARR10-B has demonstrated the residues of the α3-helix from the HTH (Helix-Turn-Helix) motif binds target DNA in the major groove while N-terminal flexible arm makes contact with the DNA minor groove (Hosoda *et al.*, 2002). It would be of great interest to map the interaction domain between AtPHR1 and AtSPX1 to find out whether the interaction with AtSPX1 interferes with dimerization of AtPHR1 and therefore affects the DNA binding, or AtSPX1 binds directly to the DNA recognition helix from the Myb domain. The recently available structural information also suggests the SPX domain undergoes a ligand binding-induced conformational change (Wild *et al.*, 2016). As SPR data shows the inhibition effect of AtSPX1 on AtdPHR1-DNA interaction is strictly phosphate-dependent, this conformational change might be of great importance in the AtSPX1-AtPHR1 interaction. The same study also showed Pi binding from yeast ScVtc2 is affected by mutations in Phosphate Binding Cluster (PBC) while InsP6 binding is controlled by both the PBC and Lysine Surface Cluster (KSC) residues (Wild *et al.*, 2016). Since mutations of AtPHO1 PBC and KSC resulted in reduced plant growth and enhanced PSI gene expression (Wild *et al.*, 2016), it would be interesting to see whether these mutations would also interfere with the AtSPX1-AtPHR1 interaction and subsequently changes the regulation of PSI genes.
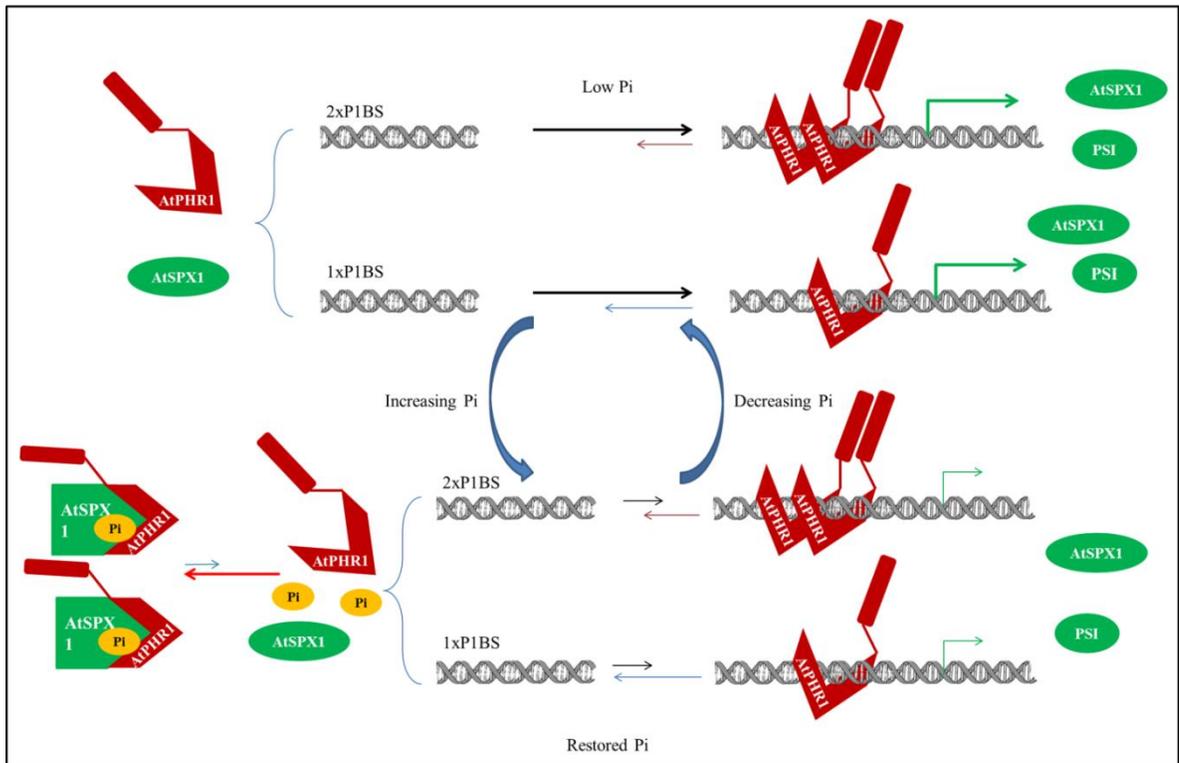
Figure 5.22 Model of how AtSPX1 might regulate PSI gene expression by influencing the AtPHR1-DNA binding equilibrium.

Under low Pi conditions, monomeric AtPHR1 in solution does not interact with AtSPX1 and the equilibrium of AtPHR1-DNA binding is in favor of AtPHR1 dimerization and DNA binding, leading to transcription of PSI genes including AtSPX1. Dissociation of AtPHR1 from DNA is slow in the absence of Pi (or IP6) and interaction with 2xP1BS motifs has a slower dissociation rate compared to 1xP1BS motifs. Under restored Pi conditions, AtSPX1 binds to Pi (or IP6) and undergoes conformational change, and is therefore able to interact with free monomeric AtPHR1 in solution. This interaction decreases the binding of AtPHR1 to DNA without changing the AtPHR1-DNA dissociation rate, therefore it deactivates the transcription of PSI genes. Genes with multiple P1BS motifs might be turned off more slowly due to a lower dissociation rate of AtPHR1. (Other components that are also involved in the expression regulation of PSI genes are not included in this sketch diagram for clarity)

178

It is of note that AtPHR1 is not only regulated by AtSPX1 through protein interaction, but is also under the sumoylation regulation of AtSIZ1. The *siz1* plants have been shown to display exaggerated Pi starvation symptoms including greater root architecture reprogramming as well as higher anthocyanin accumulation (Miura *et al.*, 2005; Miura *et al.*, 2011), indicating AtSIZ1 is a negative regulator of plant Pi starvation reaction. However, the expression level of some AtPHR1 controlled PSI genes has been down-regulated in *siz1*, suggesting the post-translational sumoylation may have a transient activation effect on AtPHR1 (Miura *et al.*, 2005). It is interesting that the two identified sumoylation sites in AtPHR1 (aa 261, 372) (Rubio *et al.*, 2001) are either within the myb DNA-binding domain between α2- and α3-helix from the HTH motif, or right after the coiled-coil domain. It is possible that the sumoylation also promotes the dimerization of AtPHR1, and the binding of AtSPX1 might inhibit the conjugation of SUMO peptide to AtPHR1.

How plants incorporate the Pi availability signal into the regulation pathways and eventually achieve the intracellular Pi homeostasis has always been a strong focus of plant phosphorus studies. SPR data from this chapter has provided a model in *Arabidopsis* where the phosphorus sensing protein AtSPX1 regulates the AtPHR1-DNA binding equilibrium by binding to monomeric transcription factor AtPHR1. This model offers a mechanistic basis for the transcriptional modulation in plant Pi regulation.

Chapter 6 General discussion and Conclusion

Prior to this study, the importance of SPX domain-containing proteins had been well established as many of these proteins were shown to play key roles in Pi regulation in eukaryotic cells (Secco *et al.*, 2012a; Secco *et al.*, 2012b). Out of the four plant SPX protein families, the SPX exclusive family members were the best characterized and proposed to function in transcriptional regulation due to their nuclear localization (Duan *et al.*, 2008). Given the vital functions of SPX proteins, and the surprisingly scarce information on their structure and interacting partners, this project sought to better characterize the role of the SPX exclusive family members in plant Pi regulation.

The first stage was to generate protein with sufficient purity and quantity to allow for structural and biochemical evaluation. To this end, throughout the project, a range of constructs containing varying sources of N-terminal SPX domains were made for expression and purification in both plant and *E. coli* host cells.

In order to obtain sufficient amount of protein material for subsequent functional and structural studies, the *Agrobacterium* mediated plant expression system was tested with the pEAQ-*HT* binary vectors which were shown to be an efficient tool in producing large quantity of recombinant proteins (Sainsbury *et al.*, 2009). Unfortunately, although the C-terminal His$_6$ tagged AtSPX1 was expressed at a suitable level, yields of purified protein were very low due to the low solubility and low stability. Similar problems were also seen from the *E. coli* expression system where most of the SPX domain constructs were successfully expressed but with very low solubility and /or poor binding to affinity resin. Denatured purification and *in vitro* refolding of insoluble SPX proteins was also attempted, although with little success. In the final 4 months of this project, a structural study on yeast and human homologues was published in *Science* which showed ligand binding helps to stabilize the otherwise disordered N-terminal α1-α2 hairpin structure of SPX domains (Wild *et al.*, 2016). This suggests specific interacting ligands and/or partners rather than specific folding mechanisms might be needed for the solubilization and stabilization of SPX proteins. In addition to ligand screening to aid in stability, over 40 different SPX domain constructs were required to achieve the structures in their study

and none of the 7 AtSPX1 constructs produced suitable material for structural studies (Wild *et al.*, 2016). The expression screening in *E. coli* host cells from this project revealed the N-terminal GST tagged AtSPX1 construct was able to produce soluble protein. The recombinant GST-AtSPX1 was recovered with relatively low purity due to poor stability, hindering the structural studies as originally planned. Although the buffer condition optimization aimed at increasing AtSPX1 stability was unsuccessful, the limited success of the GST-AtSPX1 *E. coli* expression and purification still provided sufficient material for the biochemical study of this technically challenging protein.

Contrary to the previously proposed role as transcription factors, recent studies in *Arabidopsis* and rice have demonstrated members from the SPX exclusive family can physically interact with the Myb-CC family transcription factor AtPHR1/OsPHR2 in the presence of Pi (Rubio *et al.*, 2001; Puga *et al.*, 2014; Wang *et al.*, 2014; Lv *et al.*, 2014). Using the recombinant GST-AtSPX1 and an N-terminal MBP tagged truncated version of AtPHR1 (aa 208-362), a novel mechanism for the regulation of AtPHR1 binding to P1BS by AtSPX1 was proposed and tested in this study. The *in vitro* pull-down assay using immobilized MBP-AtdPHR1 protein confirmed the Pi-dependent interaction between GST-AtSPX1 and MBP-AtdPHR1. A negative control that was lacking in the previous study (Puga *et al.*, 2014) also demonstrated this interaction is specific between AtSPX1 and AtdPHR1 as the same interaction was not observed between free GST protein and MBP-AtdPHR1, regardless of the Pi status. Although the GST protein did not interact with MBP-AtdPHR1, it bound Pi in a microscale thermophoresis experiment with a similar K_d value to GST-AtSPX1. Because of this, and the instability of the GST-AtSPX1 protein, the precise affinity of AtSPX1 for Pi could not be determined. Nevertheless, a similar K_d value for Pi binding was still obtained for GST-AtSPX1 as seen from other SPX domains (Wild *et al.*, 2016), suggesting the recombinant AtSPX1 protein was correctly folded and maintained its biological function.

In order to accurately examine how AtSPX1 regulates AtPHR1 – P1BS interaction, the associated *E. coli* host DNA needed to be removed from the recombinant transcription factor MBP-AtdPHR1. Ion-exchange chromatography was used to successfully clean up the affinity purified MBP-AtdPHR1 for subsequent DNA binding analysis. Examination of the DNA-free MBP-AtdPHR1 using SEC showed this transcription factor exists as a

monomer in solution. When interacting with DNA probes containing the specific P1BS motifs in SPR experiments, twice as much MBP-AtdPHR1 binding signal was seen from the double P1BS sites than from the single binding site. Additionally, clear protein cooperativity was also demonstrated by a much slower dissociation rate from the double binding sites. Although accurate determination of association and dissociation parameters of the MBP-AtdPHR1 – DNA interaction could not be conclusively obtained as the 1:1 model did not fit the SPR data perfectly, the stabilizing effect observed during multiple binding events still strongly suggests dimerization of MBP-AtdPHR1 protein upon DNA binding. This result is also in agreement with the EMSA data that showing AtPHR1 binds to target DNA sequences as a dimer (Rubio *et al.*, 2001). Other analytical approaches, such as analytical ultracentrifuge (AUC) or isothermal titration calorimetry (ITC), could be employed to further verify the oligomeric status of MBP-AtdPHR1 upon DNA association.

In addition to the previously observed conformational change of yeast SPX domain-containing protein Pho81 which is induced by IP7 binding, InsPs have been shown to bind SPX domains from various proteins with much higher affinities than Pi molecules (Lee *et al.*, 2007; Lee *et al.*, 2008; Wild *et al.*, 2016). However, neither 500 μM InsP6 or 5 mM Pi could promote GST-AtSPX1 to competitively displace the MBP-AtdPHR1 – DNA association in SPR experiments. This observation is contrary to the proposed model that GST-AtSPX1 displaces the MBP-AtdPHR1/P1BS interaction in the presence of Pi (Puga *et al.*, 2014). On the other hand, the GST-AtSPX1/MBP-AtdPHR1 complex formed with the same concentration of InsP6 or Pi was able to prevent the MBP-AtdPHR1 – DNA interaction, suggesting the GST-AtSPX1/MBP-AtdPHR1 is favored by the monomeric MBP-AtdPHR1 in the presence of Pi/InsPs whereas dimerization and DNA-binding could possibly bury the GST-AtSPX1 interaction site in MBP-AtdPHR1. This result provides further insight into the molecular basis of how plant integrate Pi sensing into intracellular Pi signalling cascades, demonstrating AtSPX1 regulates PSI gene expression by tuning the AtPHR1-DNA binding equilibrium under Pi sufficient conditions. The model proposed in this study (Figure 5.22) may also explain why under Pi deficient conditions, although not associated with AtPHR1, AtSPX1 is still transcriptionally up-regulated (Duan *et al.*, 2008). It is possible that the AtSPX1 is maintained at a high intracellular level so that when the Pi level is restored, it could quickly bind to the monomeric transcription factor AtPHR1 and

therefore efficiently down-regulate the expression of PSI genes under Pi sufficient conditions.

Interestingly, although the InsP6 bound K_d of ~50 μM was detected for the OsSPX2/OsPHR2 complex, InsP6 at concentration of 100 μM did not promote the interaction between GST-AtSPX1 and MBP-AtdPHR1 or the subsequent inhibitory effect on MBP-AtdPHR1 – DNA interaction. This might be the result of species-specific difference in SPX domain ligand binding, as different InsP6 binding affinities were also observed between SPX domains from *Saccharomyces cerevisiae* and *Chaetomium thermophilum* proteins (Wild *et al.*, 2016).

Although extensive functional studies of SPX domain proteins have been carried out in *Arabidopsis* and rice, the contribution of this domain to Pi regulation in other plants has not been explored extensively. Results from this study first demonstrated the involvement of SPX exclusive family members in Pi starvation responses in the agriculturally important plant potato (*Solanum tuberosum*), who's growth relies heavily on Pi nutrient.

Using a hydroponic system that was able to accurately control the nutrient input (Hammond *et al.*, 2011), morphological, physiological and gene expression responses from potato plants were investigated under different Pi conditions. Compared to the high Pi treatment (300 μM Pi input) group, a Pi deficient condition from the low Pi treatment (10 μM Pi input) group was well established by several physical responses such as increased root hair length and density, lower shoot to root ratio, as well as significantly decreased total and inorganic phosphorus content in both leaf and root tissues. Although similar total P content decrease after Pi withdrawal was seen in previous studies on potato plant, no observation on the inorganic P level changes was reported (Hammond *et al.*, 2011). P assay results from this study however demonstrated the inorganic P content decreased in the same manner during Pi starvation treatment and contributed a large part of the total P content change.

Searching the available potato genome sequence revealed 14 SPX domain-containing proteins that could be classified into the four plant SPX protein families according to the predicted C-terminal domains. The close phylogenetic relationships from these newly identified potato SPX proteins to those from *Arabidopsis* and rice showed the

183

SPX protein domains are ubiquitously found in different species and are highly conserved, suggesting conserved biological functions from this domain, such as Pi sensing (Wild *et al.*, 2016). Despite the fact that variation was observed among individual potato plants during the expression study of the SPX exclusive family corresponding genes, significant up-regulation of *StSPX2*, *StSPX3* and *StSPX5* was still seen under low Pi conditions, especially in root tissues.

Examination of the promoter regions of these genes also identified the P1BS cis-element, marking them as potential regulatory targets of Myb-CC family transcription factors, possibly homologues of AtPHR1. Additionally, multiple P1BS cis-elements were found for *StSPX2* and *StSPX5*, which were highly up-regulated during Pi starvation responses. As AtPHR1 was demonstrated to interact with P1BS motifs as a dimer (Rubio *et al.*, 2001) (Chapter 5), it is plausible that the multiple P1BS motifs from these two genes could potentially stabilize the potato transcription factor binding and therefore lead to the enhanced Pi starvation-induced expression. The fact that the P1BS cis-element is shown to be overly representative in the promoter regions of many Pi recycling-correlated genes (Acevedo-Hernández *et al.*, 2012) also suggests that it represents a useful target for designing crop plants with higher Pi use efficiency (Ruan *et al.*, 2015). AtSPX1 and AtSPX2 have been shown to negatively regulate their own expression by physically interacting with their transcription factor AtPHR1 (Puga *et al.*, 2014) (this study). Therefore, the closely related potato homologue StSPX3, which showed similar transcriptional response, might also regulate its own expression in a similar manner.

Data presented in this research shows the potato SPX-exclusive family members StSPX2, StSPX3 and StSPX5 are transcriptionally up-regulated during Pi starvation responses. Although further investigation is needed to fully understand the exact function of these proteins, the similar Pi starvation-induced expression patterns and P1BS cis-elements found in their gene promoter regions strongly suggest the existence of similar regulatory mechanisms as observed in *Arabidopsis* (Duan *et al.*, 2008; Puga *et al.*, 2014) and rice (Wang *et al.*, 2009a; Wang *et al.*, 2014). In order to investigate whether the Pi starvation-induced upregulation of certain SPX-exclusive family members is only unique to potato or generally exists in other crop plants, further identification of these SPX proteins in other plant species and investigation of their responses to Pi depletion

situations is required. It is plausible that certain SPX protein homologues, sharing very similar regulatory patterns, could represent potential targets for designing crop plants with better Pi using efficiency.

Biochemical studies on the *Arabidopsis* SPX exclusive protein AtSPX1 also demonstrated its binding activity to monomeric Myb-CC transcription factor AtPHR1 in the presence of Pi/InsP6. This interaction in turn influences the binding equilibrium of dimeric AtPHR1 and P1BS motifs and therefore negatively regulates the PSI gene expression under Pi sufficient condition. However, apart from the observations made in this project, it is interesting to notice that in the *Atspx1/Atspx2* double mutant plant, only about half of the transcriptionally altered genes are PSI genes and within these, only a small proportion are direct AtPHR1 targets (Puga *et al.*, 2014). A recent study in rice also showed that OsSPX1 is involved in anther and pollen development by regulating genes related to carbohydrate metabolism and sugar transport (Zhang *et al.*, 2016). These evidences, together with the highly plastic basic binding surface found in SPX domains (Wild *et al.*, 2016), all point to the argument that more SPX-exclusive protein interacting partners exist. In order to gain a better understanding of the full picture of plant Pi metabolism, future work is still required to identify these partners and their positions in the Pi regulation network. The *Arabidopsis* protoplast system, which was shown to successfully express and correctly localize AtSPX1, and/or the yeast two hybrid screening, may serve as good platforms for this work.

Furthermore, although the whole arrangements of different SPX domains vary due to the low similarity sequences inserted between the conserved sub-domains, the obtained structural information still provides useful insights into the future work on solving other SPX protein structures. While the solved three-dimensional structure of SPX domains is providing a powerful tool for modeling their closely related homologues (Xiang, 2006), their secondary structures will also assist in assigning domain boundaries for soluble SPX domains and determining the structures experimentally. The basic surface residues, which are highly conserved in SPX domains, suggest that they may be of high significance for the SPX domain *in vivo* functions. Mutations in these key residues may help to identify the *in vivo* function of SPX-EXS, SPX-MFS and SPX-RING family members that are currently unknown. With the help of these recent and future advances in SPX domain-

containing proteins structural and functional studies, the plant Pi sensing and signalling pathways and their strategies to cope with low Pi availability will be better understood. This will in turn help with the design of crop plants that have improved Pi use efficiency, and therefore alleviate the requirement of fertilizers and/or increase the agricultural output.

References

ACEVEDO-HERNÁNDEZ, G., OROPEZA-ABURTO, A. & HERRERA-ESTRELLA, L. 2012. A specific variant of the PHR1 binding site is highly enriched in the Arabidopsis phosphate-responsive phospholipase DZ2 coexpression network. *Plant signaling & behavior,* 7**,** 914-917.

AMES, B. N. 1966. Assay of inorganic phosphate, total phosphate and phosphatases. *In:* GINSBURG, E. N. A. V. (ed.) *Methods in Enzymology.* Academic Press, New York, NY.

ARPAT, A. B., MAGLIANO, P., WEGE, S., ROUACHED, H., STEFANOVIC, A. & POIRIER, Y. 2012. Functional expression of PHO1 to the Golgi and trans-Golgi network and its role in export of inorganic phosphate. *Plant Journal,* 71**,** 479-491.

BAKER, A., CEASAR, S. A., PALMER, A. J., PATERSON, J. B., QI, W. J., MUENCH, S. P. & BALDWIN, S. A. 2015. Replace, reuse, recycle: improving the sustainable use of phosphorus by plants. *Journal of Experimental Botany,* 66**,** 3523-3540.

BARI, R., PANT, B. D., STITT, M. & SCHEIBLE, W.-R. 2006. PHO2, microRNA399, and PHR1 define a phosphate-signaling pathway in plants. *Plant Physiology,* 141**,** 988-999.

BAYLE, V., ARRIGHI, J.-F., CREFF, A., NESPOULOUS, C., VIALARET, J., ROSSIGNOL, M., GONZALEZ, E., PAZ-ARES, J. & NUSSAUME, L. 2011. Arabidopsis thaliana high-affinity phosphate transporters exhibit multiple levels of posttranslational regulation. *Plant Cell,* 23**,** 1523-1535.

BIELESKI, R. L. 1973. Phosphate pools, phosphate transport, and phosphate availability. *Annual Review of Plant Physiology and Plant Molecular Biology,* 24**,** 225-252.

BOLANOS-GARCIA, V. M. & DAVIES, O. R. 2006. Structural analysis and classification of native proteins from E-coli commonly co-purified by immobilised metal affinity chromatography. *Biochimica Et Biophysica Acta-General Subjects,* 1760**,** 1304-1313.

BUDHIDARMO, R., NAKATANI, Y. & DAY, C. L. 2012. RINGs hold the key to ubiquitin transfer. *Trends in Biochemical Sciences,* 37**,** 58-65.

BURGESS-BROWN, N. A., SHARMA, S., SOBOTT, F., LOENARZ, C., OPPERMANN, U. & GILEADI, O. 2008. Codon optimization can improve expression of human genes in Escherichia coli: A multi-gene study. *Protein Expression and Purification,* 59**,** 94-102.

CALDERON-VAZQUEZ, C., IBARRA-LACLETTE, E., CABALLERO-PEREZ, J. & HERRERA-ESTRELLA, L. 2008. Transcript profiling of Zea mays roots reveals gene responses to phosphate deficiency at the plant- and species-specific levels. *Journal of Experimental Botany,* 59**,** 2479-2497.

CHAN, F. Y. & TORRIANI, A. 1996. PstB protein of the phosphate-specific transport system of Escherichia coli is an ATPase. *Journal of Bacteriology,* 178**,** 3974-3977.

CHENG, Y. X., ZHOU, W. B., EL SHEERY, N. I., PETERS, C., LI, M. Y., WANG, X. M. & HUANG, J. R. 2011. Characterization of the Arabidopsis glycerophosphodiester phosphodiesterase (GDPD) family reveals a role of the plastid-localized AtGDPD1 in maintaining cellular phosphate homeostasis under phosphate starvation. *Plant Journal,* 66**,** 781-795.

CHEVALIER, F., PATA, M., NACRY, P., DOUMAS, P. & ROSSIGNOL, M. 2003. Effects of phosphate availability on the root system architecture: large-scale analysis of the natural variation between Arabidopsis accessions. *Plant Cell and Environment,* 26**,** 1839-1850.

CORDELL, D., DRANGERT, J. O. & WHITE, S. 2009. The story of phosphorus: Global food security and food for thought. *Global Environmental Change-Human and Policy Dimensions,* 19**,** 292-305.

DELHAIZE, E. & RANDALL, P. J. 1995. Characterization of a phosphate-accumulator mutant of Arabidopsis thaliana. *Plant Physiology,* 107**,** 207-213.

DESAI, M., RANGARAJAN, P., DONAHUE, J. L., WILLIAMS, S. P., LAND, E. S., MANDAL, M. K., PHILLIPPY, B. Q., PERERA, I. Y., RABOY, V. & GILLASPY, G. E. 2014. Two inositol hexakisphosphate kinases drive inositol pyrophosphate synthesis in plants. *Plant Journal,* 80**,** 642-653.

DEVAIAH, B. N., KARTHIKEYAN, A. S. & RAGHOTHAMA, K. G. 2007. WRKY75 transcription factor is a modulator of phosphate acquisition and root development in arabidopsis. *Plant Physiology,* 143**,** 1789-1801.

DEVAIAH, B. N., MADHUVANTHI, R., KARTHIKEYAN, A. S. & RAGHOTHAMA, K. G. 2009. Phosphate starvation responses and gibberellic acid biosynthesis are regulated by the MYB62 transcription factor in Arabidopsis. *Molecular Plant,* 2**,** 43-58.

DUAN, K., YI, K., DANG, L., HUANG, H., WU, W. & WU, P. 2008. Characterization of a sub-family of Arabidopsis genes with the SPX domain reveals their diverse functions in plant tolerance to phosphorus starvation. *Plant Journal,* 54**,** 965-975.

DUBOS, C., STRACKE, R., GROTEWOLD, E., WEISSHAAR, B., MARTIN, C. & LEPINIEC, L. 2010. MYB transcription factors in Arabidopsis. *Trends in Plant Science,* 15**,** 573-581.

DUNLOP, J., PHUNG, H. T., MEEKING, R. & WHITE, D. W. R. 1997. The kinetics associated with phosphate absorption by Arabidopsis and its regulation by phosphorus status. *Australian Journal of Plant Physiology,* 24**,** 623-629.

DYSON, M. R. 2010. Selection of soluble protein expression constructs: the experimental determination of protein domain boundaries. *Biochemical Society Transactions,* 38**,** 908-913.

EDELHEIT, O., HANUKOGLU, A. & HANUKOGLU, I. 2009. Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *Bmc Biotechnology,* 9**,** 61.

ELSER, J. & BENNETT, E. 2011. A broken biogeochemical cycle. *Nature,* 478**,** 29-31.

ENDLER, A., MEYER, S., SCHELBERT, S., SCHNEIDER, T., WESCHKE, W., PETERS, S. W., KELLER, F., BAGINSKY, S., MARTINOIA, E. & SCHMIDT, U. G. 2006. Identification of a vacuolar sucrose transporter in barley and arabidopsis mesophyll cells by a tonoplast proteomic approach. *Plant Physiology,* 141**,** 196-207.

FISHER, E., ALMAGUER, C., HOLIC, R., GRIAC, P. & PATTON-VOGT, J. 2005. Glycerophosphocholine-dependent growth requires Gde1p (YPL110c) and Git1p in Saccharomyces cerevisiae. *Journal of Biological Chemistry,* 280**,** 36110-36117.

GARDNER, S. G., JOHNS, K. D., TANNER, R. & MCCLEARY, W. R. 2014. The PhoU protein from Escherichia coli interacts with PhoR, PstB, and metals to form a phosphate-signaling complex at the membrane. *Journal of Bacteriology,* 196**,** 1741-1752.

GHILLEBERT, R., SWINNEN, E., DE SNIJDER, P., SMETS, B. & WINDERICKX, J. 2011. Differential roles for the low-affinity phosphate transporters Pho87 and Pho90 in Saccharomyces cerevisiae. *Biochemical Journal,* 434**,** 243-251.

GIOTS, F., DONATON, M. C. V. & THEVELEIN, J. M. 2003. Inorganic phosphate is sensed by specific phosphate carriers and acts in concert with glucose as a nutrient signal for activation of the protein kinase A pathway in the yeast Saccharomyces cerevisiae. *Molecular Microbiology,* 47**,** 1163-1181.

GIOVANNINI, D., TOUHAMI, J., CHARNET, P., SITBON, M. & BATTINI, J. L. 2013. Inorganic phosphate export by the retrovirus receptor XPR1 in metazoans. *Cell Reports,* 3**,** 1866-1873.

GONZÁLEZ, E., SOLANO, R., RUBIO, V., LEYVA, A. & PAZ-ARES, J. 2005. PHOSPHATE TRANSPORTER TRAFFIC FACILITATOR1 is a plant-specific SEC12-related protein that enables the endoplasmic reticulum exit of a high-affinity phosphate transporter in Arabidopsis. *Plant Cell,* 17**,** 3500-3512.

GORDON-WEEKS, R., TONG, Y. P., DAVIES, T. G. E. & LEGGEWIE, G. 2003. Restricted spatial expression of a high-affinity phosphate transporter in potato roots. *Journal of Cell Science,* 116**,** 3135-3144.

GRAEBER, K., LINKIES, A., WOOD, A. T. A. & LEUBNER-METZGER, G. 2011. A guideline to family-wide comparative state-of-the-art quantitative RT-PCR analysis exemplified with a Brassicaceae cross-species seed germination case study. *Plant Cell,* 23**,** 2045-2063.

GRILLO-PUERTAS, M., SCHURIG-BRICCIO, L. A., RODRIGUEZ-MONTELONGO, L., RINTOUL, M. R. & RAPISARDA, V. A. 2014. Copper tolerance mediated by polyphosphate degradation and low-affinity inorganic phosphate transport system in Escherichia coli. *Bmc Microbiology,* 14**,** 72.

GROTE, A., HILLER, K., SCHEER, M., MUNCH, R., NORTEMANN, B., HEMPEL, D. C. & JAHN, D. 2005. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Research,* 33**,** W526-W531.

HAMBURGER, D., REZZONICO, E., PETETOT, J. M. C., SOMERVILLE, C. & POIRIER, Y. 2002. Identification and characterization of the Arabidopsis PHO1 gene involved in phosphate loading to the xylem. *Plant Cell,* 14**,** 889-902.

HAMMOND, J. P., BENNETT, M. J., BOWEN, H. C., BROADLEY, M. R., EASTWOOD, D. C., MAY, S. T., RAHN, C., SWARUP, R., WOOLAWAY, K. E. & WHITE, P. J. 2003. Changes in gene expression in Arabidopsis shoots during phosphate starvation and the potential for developing smart plants. *Plant Physiology,* 132**,** 578-596.

HAMMOND, J. P., BROADLEY, M. R., BOWEN, H. C., SPRACKLEN, W. P., HAYDEN, R. M. & WHITE, P. J. 2011. Gene expression changes in phosphorus deficient potato (Solanum tuberosum L.) leaves and the potential for diagnostic gene expression markers. *Plos One,* 6**,** e24606.

HAMMOND, J. P. & WHITE, P. J. 2011. Sugar signaling in root responses to low phosphorus availability. *Plant Physiology,* 156**,** 1033-1040.

HARRISON, M. J. 2005. Signaling in the arbuscular mycorrhizal symbiosis. *Annual Review of Microbiology,* 59**,** 19-42.

HIGGINS, D. G., THOMPSON, J. D. & GIBSON, T. J. 1996. Using CLUSTAL for multiple sequence alignments. *Computer Methods for Macromolecular Sequence Analysis,* 266**,** 383-402.

HOLMES, L. 2012. British Survey of Fertiliser Practice: fertiliser use on farm crops for crop year 2011.: DEFRA. ISBN 978-0-95525-697-4. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/181432/defra-stats-foodfarm-environ-fertiliserpractice-2011-120425.pdf. Accessed on 18/01/2014.

HOLSBEEKS, I., LAGATIE, O., VAN NULAND, A., VAN DE VELDE, S. & THEVELEIN, J. M. 2004. The eukaryotic plasma membrane as a nutrient-sensing device. *Trends in Biochemical Sciences,* 29**,** 556-564.

HOSODA, K., IMAMURA, A., KATOH, E., HATTA, T., TACHIKI, M., YAMADA, H., MIZUNO, T. & YAMAZAKI, T. 2002. Molecular structure of the GARP family of plant Myb-related DNA binding motifs of the Arabidopsis response regulators. *Plant Cell,* 14**,** 2015-2029.

HOTHORN, M., NEUMANN, H., LENHERR, E. D., WEHNER, M., RYBIN, V., HASSA, P. O., UTTENWEILER, A., REINHARDT, M., SCHMIDT, A., SEILER, J., LADURNER, A. G., HERRMANN, C., SCHEFFZEK, K. & MAYER, A. 2009. Catalytic core of a membrane-associated eukaryotic polyphosphate polymerase. *Science,* 324**,** 513-516.

HSIEH, L. C., LIN, S. I., SHIH, A. C. C., CHEN, J. W., LIN, W. Y., TSENG, C. Y., LI, W. H. & CHIOU, T. J. 2009. Uncovering small RNA-mediated responses to phosphate deficiency in Arabidopsis by deep sequencing. *Plant Physiology,* 151**,** 2120-2132.

HSIEH, Y. J. & WANNER, B. L. 2010. Global regulation by the seven-component Pi signaling system. *Current Opinion in Microbiology,* 13**,** 198-203.

HÜRLIMANN, H. C., PINSON, B., STADLER-WAIBEL, M., ZEEMAN, S. C. & FREIMOSER, F. M. 2009. The SPX domain of the yeast low-affinity phosphate transporter Pho90 regulates transport activity. *Embo Reports,* 10**,** 1003-1008.

HÜRLIMANN, H. C., STADLER-WAIBEL, M., WERNER, T. P. & FREIMOSER, F. M. 2007. Pho91 is a vacuolar phosphate transporter that regulates phosphate and polyphosphate metabolism in Saccharomyces cerevisiae. *Molecular biology of the cell,* 18**,** 4438-4445.

JACKSON, R. J., BINET, M. R. B., LEE, L. J., MA, R., GRAHAM, A. I., MCLEOD, C. W. & POOLE, R. K. 2008. Expression of the PitA phosphate/metal transporter of Escherichia coli is responsive to zinc and inorganic phosphate levels. *Fems Microbiology Letters,* 289**,** 219-224.

JAVOT, H., PENMETSA, R. V., TERZAGHI, N., COOK, D. R. & HARRISON, M. J. 2007. A Medicago truncatula phosphate transporter indispensable for the arbuscular mycorrhizal symbiosis. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 1720-1725.

JIAO, W. T., CHEN, W. P., CHANG, A. C. & PAGE, A. L. 2012. Environmental risks of trace elements associated with long-term phosphate fertilizers applications: A review. *Environmental Pollution,* 168**,** 44-53.

KANT, S., PENG, M. & ROTHSTEIN, S. J. 2011. Genetic regulation by NLA and microRNA827 for maintaining nitrate-dependent phosphate homeostasis in Arabidopsis. *Plos Genetics,* 7**,** e1002021.

KAPUST, R. B. & WAUGH, D. S. 1999. Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Science,* 8**,** 1668-1674.

KARTHIKEYAN, A. S., BALLACHANDA, D. N. & RAGHOTHAMA, K. G. 2009. Promoter deletion analysis elucidates the role of cis elements and 5'UTR intron in spatiotemporal regulation of AtPht1;4 expression in Arabidopsis. *Physiologia Plantarum,* 136**,** 10-18.

KHOW, O. & SUNTRARACHUN, S. 2012. Strategies for production of active eukaryotic proteins in bacterial expression system. *Asian Pacific Journal of Tropical Biomedicine,* 2**,** 159-162.

KIM, B. R., NAM, H. Y., KIM, S. U., KIM, S. I. & CHANG, Y. J. 2003. Normalization of reverse transcription quantitative-PCR with housekeeping genes in rice. *Biotechnology Letters,* 25**,** 1869-1872.

KOMEILI, A. & O'SHEA, E. K. 1999. Roles of phosphorylation sites in regulating activity of the transcription factor Pho4. *Science,* 284**,** 977-980.

LEE, Y.-S., MULUGU, S., YORK, J. D. & O'SHEA, E. K. 2007. Regulation of a cyclin-CDK-CDK inhibitor complex by inositol pyrophosphates. *Science,* 316**,** 109-112.

LEE, Y. S., HUANG, K. X., QUIOCHO, F. A. & O'SHEA, E. K. 2008. Molecular basis of cyclin-CDK-CKI regulation by reversible binding of an inositol pyrophosphate. *Nature Chemical Biology,* 4**,** 25-32.

LEGGEWIE, G., WILLMITZER, L. & RIESMEIER, J. W. 1997. Two cDNAs from potato are able to complement a phosphate uptake-deficient yeast mutant: Identification of phosphate transporters from higher plants. *Plant Cell,* 9**,** 381-392.

LIBANATI, C. M. & TANDLER, C. J. 1969. The distribution of the water-soluble inorganic orthophosphate ions within the cell: accumulation in the nucleus. Electron probe microanalysis. *Journal of Cell Biology,* 42**,** 754-765.

LIN, S.-I., SANTI, C., JOBET, E., LACUT, E., EL KHOLTI, N., KARLOWSKI, W. M., VERDEIL, J.-L., BREITLER, J. C., PERIN, C., KO, S.-S., GUIDERDONI, E., CHIOU, T.-J. & ECHEVERRIA, M. 2010. Complex regulation of two target genes encoding SPX-MFS proteins by rice miR827 in response to phosphate starvation. *Plant and Cell Physiology,* 51**,** 2119-2131.

LIN, W. Y., HUANG, T. K. & CHIOU, T. J. 2013. Nitrogen limitation adaptation, a target of microRNA827, mediates degradation of plasma membrane-localized phosphate transporters to maintain phosphate homeostasis in Arabidopsis. *Plant Cell,* 25**,** 4061-4074.

LIN, W. Y., LIN, S. I. & CHIOU, T. J. 2009. Molecular regulators of phosphate homeostasis in plants. *Journal of Experimental Botany,* 60**,** 1427-1438.

LIU, C. M., MUCHHAL, U. S., UTHAPPA, M., KONONOWICZ, A. K. & RAGHOTHAMA, K. G. 1998. Tomato phosphate transporter genes are differentially regulated in plant tissues by phosphorus. *Plant Physiology,* 116**,** 91-99.

LIU, J., YANG, L., LUAN, M., WANG, Y., ZHANG, C., ZHANG, B., SHI, J., ZHAO, F.-G., LAN, W. & LUAN, S. 2015. A vacuolar phosphate transporter essential for phosphate homeostasis in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America,* 112**,** E6571-E6578.

LIU, T. Y., HUANG, T. K., TSENG, C. Y., LAI, Y. S., LIN, S. I., LIN, W. Y., CHEN, J. W. & CHIOU, T. J. 2012. PHO2-dependent degradation of PHO1 modulates phosphate homeostasis in Arabidopsis. *Plant Cell,* 24**,** 2168-2183.

LIVAK, K. J. & SCHMITTGEN, T. D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods,* 25**,** 402-408.

LLOYD, J. C. & ZAKHLENIUK, O. V. 2004. Responses of primary and secondary metabolism to sugar accumulation revealed by microarray expression analysis of the Arabidopsis mutant, pho3. *Journal of Experimental Botany,* 55**,** 1221-1230.

LÓPEZ-ARREDONDO, D. L., LEYVA-GONZALEZ, M. A., GONZALEZ-MORALES, S. I., LOPEZ-BUCIO, J. & HERRERA-ESTRELLA, L. 2014. Phosphate nutrition: improving low-phosphate tolerance in crops. *Annual Review of Plant Biology, Vol 65,* 65**,** 95-123.

LOUGHEED, T. 2011. Phosphorus paradox: scarcity and overabundance of a key nutrient. *Environmental health perspectives,* 119**,** A208-A213.

LUBIN, E. A., HENRY, J. T., FIEBIG, A., CROSSON, S. & LAUB, M. T. 2016. Identification of the PhoB regulon and role of PhoU in the phosphate starvation response of Caulobacter crescentus. *Journal of Bacteriology,* 198**,** 187-200.

LUNDH, F., MOUILLON, J. M., SAMYN, D., STADLER, K., POPOVA, Y., LAGERSTEDT, J. O., THEVELEIN, J. M. & PERSSON, B. L. 2009. Molecular mechanisms controlling phosphate-induced downregulation of the yeast Pho84 phosphate transporter. *Biochemistry,* 48**,** 4497-4505.

LV, Q., ZHONG, Y., WANG, Y., ZHANG, L., SHI, J., WU, Z., LIU, Y., MAO, C., YI, K. & WU, P. 2014. SPX4 negatively regulates phosphate signaling and homeostasis through its interaction with PHR2 in rice. *Plant Cell,* 26**,** 1586-1597.

MARSCHNER, H. 2012. *Marschner's Mineral Nutrition of Higher Plants, 3rd Edition,* San Diego, London, Academic Press.

MATIĆ, S., MASENGA, V., POLI, A., RINALDI, R., MILNE, R. G., VECCHIATI, M. & NORIS, E. 2012. Comparative analysis of recombinant Human Papillomavirus 8 L1

production in plants by a variety of expression systems and purification methods. *Plant Biotechnology Journal,* 10**,** 410-421.

MATIĆ, S., RINALDI, R., MASENGA, V. & NORIS, E. 2011. Efficient production of chimeric Human papillomavirus 16 L1 protein bearing the M2e influenza epitope in Nicotiana benthamiana plants. *Bmc Biotechnology,* 11**,** 106.

MAYER, M. P. & BUKAU, B. 2005. Hsp70 chaperones: Cellular functions and molecular mechanism. *Cellular and Molecular Life Sciences,* 62**,** 670-684.

MIROUX, B. & WALKER, J. E. 1996. Over-production of proteins in Escherichia coli: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology,* 260**,** 289-298.

MISSON, J., THIBAUD, M. C., BECHTOLD, N., RAGHOTHAMA, K. & NUSSAUME, L. 2004. Transcriptional regulation and functional properties of Arabidopsis Pht1;4, a high affinity transporter contributing greatly to phosphate uptake in phosphate deprived plants. *Plant Molecular Biology,* 55**,** 727-741.

MITSUKAWA, N., OKUMURA, S., SHIRANO, Y., SATO, S., KATO, T., HARASHIMA, S. & SHIBATA, D. 1997. Overexpression of an Arabidopsis thaliana high-affinity phosphate transporter gene in tobacco cultured cells enhances cell growth under phosphate-limited conditions. *Proceedings of the National Academy of Sciences of the United States of America,* 94**,** 7098-7102.

MIURA, K., LEE, J., GONG, Q. Q., MA, S. S., JIN, J. B., YOO, C. Y., MIURA, T., SATO, A., BOHNERT, H. J. & HASEGAWA, P. M. 2011. SIZ1 regulation of phosphate starvation-induced root architecture remodeling involves the control of auxin accumulation. *Plant Physiology,* 155**,** 1000-1012.

MIURA, K., RUS, A., SHARKHUU, A., YOKOI, S., KARTHIKEYAN, A. S., RAGHOTHAMA, K. G., BAEK, D., KOO, Y. D., JIN, J. B., BRESSAN, R. A., YUN, D. J. & HASEGAWA, P. M. 2005. The Arabidopsis SUMO E3 ligase SIZ1 controls phosphate deficiency responses. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 7760-7765.

MUCHHAL, U. S., PARDO, J. M. & RAGHOTHAMA, K. G. 1996. Phosphate transporters from the higher plant Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America,* 93**,** 10519-10523.

MUDGE, S. R., RAE, A. L., DIATLOFF, E. & SMITH, F. W. 2002. Expression analysis suggests novel roles for members of the Pht1 family of phosphate transporters in Arabidopsis. *Plant Journal,* 31**,** 341-353.

MÜLLER, J., GODDE, V., NIEHAUS, K. & ZORB, C. 2015. Metabolic adaptations of White Lupin roots and shoots under phosphorus deficiency. *Frontiers in Plant Science,* 6**,** 1014.

NGUYEN, M. P., YOON, J. M., CHO, M. H. & LEE, S. W. 2015. Prokaryotic 2-component systems and the OmpR/PhoB superfamily. *Canadian Journal of Microbiology,* 61**,** 799-810.

NICOT, N., HAUSMAN, J. F., HOFFMANN, L. & EVERS, D. 2005. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *Journal of Experimental Botany,* 56**,** 2907-2914.

NUSSAUME, L., KANNO, S., JAVOT, H., MARIN, E., POCHON, N., AYADI, A., NAKANISHI, T. M. & THIBAUD, M.-C. 2011. Phosphate import in plants: focus on the PHT1 transporters. *Frontiers in Plant Science,* 2**,** 83.

OGANESYAN, V., OGANESYAN, N., ADAMS, P. D., JANCARIK, J., YOKOTA, H. A., KIM, R. & KIM, S. H. 2005. Crystal structure of the "PhoU-Like" phosphate uptake regulator from Aquifex aeolicus. *Journal of Bacteriology,* 187**,** 4238-4244.

OGAWA, N., DERISI, J. & BROWN, P. O. 2000. New components of a system for phosphate accumulation and polyphosphate metabolism in Saccharomyces cerevisiae revealed by genomic expression analysis. *Molecular Biology of the Cell,* 11**,** 4309-4321.

OHSHIMA, N., YAMASHITA, S., TAKAHASHI, N. & KUROISHI, C. 2008. Escherichia coli cytosolic glycerophosphodiester phosphodiesterase (UgpQ) requires Mg^{2+}, Co^{2+}, or Mn^{2+} for its enzyme activity. *Journal of Bacteriology,* 190**,** 1219-1223.

OROPEZA-ABURTO, A., CRUZ-RAMIREZ, A., ACEVEDO-HERNANDEZ, G. J., PEREZ-TORRES, C.-A., CABALLERO-PEREZ, J. & HERRERA-ESTRELLA, L. 2012. Functional analysis of the Arabidopsis PLDZ2 promoter reveals an evolutionarily conserved low-Pi-responsive transcriptional enhancer element. *Journal of Experimental Botany,* 63**,** 2189-2202.

PAGE, R. D. M. 1996. TreeView: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences,* 12**,** 357-358.

PAO, S. S., PAULSEN, I. T. & SAIER, M. H. 1998. Major facilitator superfamily. *Microbiology and Molecular Biology Reviews,* 62**,** 1-34.

PARNISKE, M. 2004. Molecular genetics of the arbuscular mycorrhizal symbiosis. *Current Opinion in Plant Biology,* 7**,** 414-421.

PENG, M., HANNAM, C., GU, H., BI, Y.-M. & ROTHSTEIN, S. J. 2007. A mutation in NLA, which encodes a RING-type ubiquitin ligase, disrupts the adaptability of Arabidopsis to nitrogen limitation. *Plant Journal,* 50**,** 320-337.

POIRIER, Y., THOMA, S., SOMERVILLE, C. & SCHIEFELBEIN, J. 1991. Mutant of Arabidopsis deficient in xylem loading of phosphate. *Plant Physiology,* 97**,** 1087-1093.

PRATT, J., BOISSON, A.-M., GOUT, E., BLIGNY, R., DOUCE, R. & AUBERT, S. 2009. Phosphate (Pi) starvation effect on the cytosolic Pi concentration and Pi exchanges across the tonoplast in plant cells: an in vivo 31P-nuclear magnetic resonance study using methylphosphonate as a Pi analog. *Plant Physiology,* 151**,** 1646-1657.

PUGA, M. I., MATEOS, I., CHARUKESI, R., WANG, Z., FRANCO-ZORRILLA, J. M., DE LORENZO, L., IRIGOYE, M. L., MASIERO, S., BUSTOS, R., RODRIGUEZ, J., LEYVA, A., RUBIO, V., SOMMER, H. & PAZ-ARES, J. 2014. SPX1 is a phosphate-dependent inhibitor of Phosphate Starvation Response 1 in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America,* 111**,** 14947-14952.

PUMPLIN, N., ZHANG, X. C., NOAR, R. D. & HARRISON, M. J. 2012. Polar localization of a symbiosis-specific phosphate transporter is mediated by a transient reorientation of secretion. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** E665-E672.

QI, W., BALDWIN, S. A., MUENCH, S. P. & BAKER, A. 2016. Pi sensing and signalling: from prokaryotic to eukaryotic cells. *Biochemical Society Transactions,* 44**,** 766-773.

RADONIC, A., THULKE, S., MACKAY, I. M., LANDT, O., SIEGERT, W. & NITSCHE, A. 2004. Guideline to reference gene selection for quantitative real-time PCR. *Biochemical and Biophysical Research Communications,* 313**,** 856-862.

RAE, A. L., CYBINSKI, D. H., JARMEY, J. M. & SMITH, F. W. 2003. Characterization of two phosphate transporters from barley; evidence for diverse function and kinetic properties among members of the Pht1 family. *Plant Molecular Biology,* 53**,** 27-36.

RAGHOTHAMA, K. G. & KARTHIKEYAN, A. S. 2005. Phosphate acquisition. *Plant and Soil,* 274**,** 37-49.

RARAN-KURUSSI, S., KEEFE, K. & WAUGH, D. S. 2015. Positional effects of fusion partners on the yield and solubility of MBP fusion proteins. *Protein Expression and Purification,* 110**,** 159-164.

REYMOND, M., SVISTOONOFF, S., LOUDET, O., NUSSAUME, L. & DESNOS, T. 2006. Identification of QTL controlling root growth response to phosphate starvation in Arabidopsis thaliana. *Plant Cell and Environment,* 29**,** 115-125.

RIBOT, C., ZIMMERLI, C., FARMER, E. E., REYMOND, P. & POIRIER, Y. 2008. Induction of the Arabidopsis PHO1;H10 gene by 12-oxo-phytodienoic acid but not jasmonic acid via a CORONATINE INSENSITIVE1-dependent pathway. *Plant Physiology,* 147**,** 696-706.

ROELAND BOER, D., FREIRE-RIOS, A., VAN DEN BERG, W. A. M., SAAKI, T., MANFIELD, I. W., KEPINSKI, S., LOPEZ-VIDRIEO, I., MANUEL FRANCO-ZORRILLA, J., DE VRIES, S. C., SOLANO, R., WEIJERS, D. & COLL, M. 2014. Structural basis for DNA binding specificity by the auxin-dependent ARF transcription factors. *Cell,* 156**,** 577-589.

ROSANO, G. L. & CECCARELLI, E. A. 2009. Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted Escherichia coli strain. *Microbial Cell Factories,* 8**,** 41.

RUAN, W. Y., GUO, M. N., CAI, L. L., HU, H. T., LI, C. Y., LIU, Y., WU, Z. C., MAO, C. Z., YI, K. K., WU, P. & MO, X. R. 2015. Genetic manipulation of a high-affinity PHR1 target cis-elementto improve phosphorous uptake in Oryza sativa L. *Plant Molecular Biology,* 87**,** 429-440.

RUBIO, V., LINHARES, F., SOLANO, R., MARTIN, A. C., IGLESIAS, J., LEYVA, A. & PAZ-ARES, J. 2001. A conserved MYB transcription factor involved in phosphate starvation signaling both in vascular plants and in unicellular algae. *Genes & Development,* 15**,** 2122-2133.

RUDOLPH, R. & LILIE, H. 1996. In vitro folding of inclusion body proteins. *Faseb Journal,* 10**,** 49-56.

RUIJTER, J. M., PFAFFL, M. W., ZHAO, S., SPIESS, A. N., BOGGY, G., BLOM, J., RUTLEDGE, R. G., SISTI, D., LIEVENS, A., DE PRETER, K., DERVEAUX, S., HELLEMANS, J. & VANDESOMPELE, J. 2013. Evaluation of qPCR curve analysis methods for reliable biomarker discovery: Bias, resolution, precision, and implications. *Methods,* 59**,** 32-46.

RUTLEDGE, R. G. & CÔTÉ, C. 2003. Mathematics of quantitative kinetic PCR and the application of standard curves. *Nucleic Acids Research,* 31**,** e93.

SAINSBURY, F., THUENEMANN, E. C. & LOMONOSSOFF, G. P. 2009. pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnology Journal,* 7**,** 682-693.

SÁNCHEZ-CALDERÓN, L., LOPEZ-BUCIO, J., CHACON-LOPEZ, A., CRUZ-RAMIREZ, A., NIETO-JACOBO, F., DUBROVSKY, J. G. & HERRERA-ESTRELLA, L. 2005. Phosphate starvation induces a determinate developmental program in the roots of Arabidopsis thaliana. *Plant and Cell Physiology,* 46**,** 174-184.

SATO, A. & MIURA, K. 2011. Root architecture remodeling induced by phosphate starvation. *Plant signaling & behavior,* 6**,** 1122-1126.

SCHACHTMAN, D. P., REID, R. J. & AYLING, S. M. 1998. Phosphorus uptake by plants: From soil to cell. *Plant Physiology,* 116**,** 447-453.

SECCO, D., BAUMANN, A. & POIRIER, Y. 2010. Characterization of the rice PHO1 gene family reveals a key role for OsPHO1;2 in phosphate homeostasis and the evolution of a distinct clade in dicotyledons. *Plant Physiology,* 152**,** 1693-1704.

SECCO, D., WANG, C., ARPAT, B. A., WANG, Z., POIRIER, Y., TYERMAN, S. D., WU, P., SHOU, H. & WHELAN, J. 2012a. The emerging importance of the SPX domain-containing proteins in phosphate homeostasis. *New Phytologist,* 193**,** 842-851.

SECCO, D., WANG, C., SHOU, H. X. & WHELAN, J. 2012b. Phosphate homeostasis in the yeast Saccharomyces cerevisiae, the key role of the SPX domain-containing proteins. *Febs Letters,* 586**,** 289-295.

SHIN, H., SHIN, H. S., DEWBRE, G. R. & HARRISON, M. J. 2004. Phosphate transport in Arabidopsis: Pht1;1 and Pht1;4 play a major role in phosphate acquisition from both low- and high-phosphate environments. *Plant Journal,* 39**,** 629-642.

SOK, D. E. 1998. Active site of brain Zn^{2+}-glycerophosphocholine cholinephosphodiesterase and regulation of enzyme activity. *Neurochemical Research,* 23**,** 1061-1067.

SPARKES, I. A., RUNIONS, J., KEARNS, A. & HAWES, C. 2006. Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nature Protocols,* 1**,** 2019-2025.

STEFANOVIC, A., ARPAT, A. B., BLIGNY, R., GOUT, E., VIDOUDEZ, C., BENSIMON, M. & POIRIER, Y. 2011. Over-expression of PHO1 in Arabidopsis leaves reveals its role in mediating phosphate efflux. *Plant Journal,* 66**,** 689-699.

STEFANOVIC, A., RIBOT, C., ROUACHED, H., WANG, Y., CHONG, J., BELBAHRI, L., DELESSERT, S. & POIRIER, Y. 2007. Members of the PHO1 gene family show limited functional redundancy in phosphate transfer to the shoot, and are regulated by phosphate deficiency via distinct pathways. *Plant Journal,* 50**,** 982-994.

STUDIER, F. W. 2005. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification,* 41**,** 207-234.

STÜRZENBAUM, S. R. & KILLE, P. 2001. Control genes in quantitative molecular biological techniques: the variability of invariance. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology,* 130**,** 281-289.

TELLINGHUISEN, J. & SPIESS, A.-N. 2014. Comparing real-time quantitative polymerase chain reaction analysis methods for precision, linearity, and accuracy of estimating amplification efficiency. *Analytical Biochemistry,* 449**,** 76-82.

THOMPSON, J. D., GIBSON, T. J., PLEWNIAK, F., JEANMOUGIN, F. & HIGGINS, D. G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research,* 25**,** 4876-4882.

TICCONI, C. A., DELATORRE, C. A., LAHNER, B., SALT, D. E. & ABEL, S. 2004. Arabidopsis pdr2 reveals a phosphate-sensitive checkpoint in root development. *Plant Journal,* 37**,** 801-814.

TICCONI, C. A., LUCERO, R. D., SAKHONWASEE, S., ADAMSON, A. W., CREFF, A., NUSSAUME, L., DESNOS, T. & ABEL, S. 2009. ER-resident proteins PDR2 and LPR1 mediate the developmental response of root meristems to phosphate availability. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 14174-14179.

VALLEJO, L. F. & RINAS, U. 2004. Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. *Microbial Cell Factories,* 3**,** 11.

VAN DER REST, B., ROLLAND, N., BOISSON, A. M., FERRO, M., BLIGNY, R. & DOUCE, R. 2004. Identification and characterization of plant glycerophosphodiester phosphodiesterase. *Biochemical Journal,* 379**,** 601-607.

VOLKOV, R. A., PANCHUK, II & SCHOFFL, F. 2003. Heat-stress-dependency and developmental modulation of gene expression: the potential of house-keeping genes as internal standards in mRNA expression profiling using real-time RT-PCR. *Journal of Experimental Botany,* 54**,** 2343-2349.

WAGNER, S., KLEPSCH, M. M., SCHLEGEL, S., APPEL, A., DRAHEIM, R., TARRY, M., HOGBOM, M., VAN WIJK, K. J., SLOTBOOM, D. J., PERSSON, J. O. & DE GIER, J. W. 2008. Tuning Escherichia coli for membrane protein overexpression. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 14371-14376.

WANG, C., HUANG, W., YING, Y., LI, S., SECCO, D., TYERMAN, S., WHELAN, J. & SHOU, H. 2012a. Functional characterization of the rice SPX-MFS family reveals a key role of OsSPX-MFS1 in controlling phosphate homeostasis in leaves. *New Phytologist,* 196**,** 139-148.

WANG, C., YING, S., HUANG, H. J., LI, K., WU, P. & SHOU, H. X. 2009a. Involvement of OsSPX1 in phosphate homeostasis in rice. *Plant Journal,* 57**,** 895-904.

WANG, L. S., DONG, J. S., GAO, Z. Y. & LIU, D. 2012b. The Arabidopsis gene HYPERSENSITIVE TO PHOSPHATE STARVATION 3 encodes ETHYLENE OVERPRODUCTION 1. *Plant and Cell Physiology,* 53**,** 1093-1105.

WANG, Y., RIBOT, C., REZZONICO, E. & POIRIER, Y. 2004. Structure and expression profile of the Arabidopsis PHO1 gene family indicates a broad role in inorganic phosphate homeostasis. *Plant Physiology,* 135**,** 400-411.

WANG, Z., HU, H., HUANG, H., DUAN, K., WU, Z. & WU, P. 2009b. Regulation of OsSPX1 and OsSPX3 on Expression of OsSPX domain Genes and Pi-starvation Signaling in Rice. *Journal of Integrative Plant Biology,* 51**,** 663-674.

WANG, Z., RUAN, W., SHI, J., ZHANG, L., XIANG, D., YANG, C., LI, C., WU, Z., LIU, Y., YU, Y., SHOU, H., MO, X., MAO, C. & WU, P. 2014. Rice SPX1 and SPX2 inhibit phosphate starvation responses through interacting with PHR2 in a phosphate-dependent manner. *Proceedings of the National Academy of Sciences of the United States of America,* 111**,** 14953-14958.

WANNER, B. L. 1996. Signal transduction in the control of phosphate-regulated genes of Escherichia coli. *Kidney International,* 49**,** 964-967.

WASAKI, J., YONETANI, R., SHINANO, T., KAI, M. & OSAKI, M. 2003. Expression of the OsPI1 gene, cloned from rice roots using cDNA microarray, rapidly responds to phosphorus status. *New Phytologist,* 158**,** 239-248.

WAUGH, D. S. 2005. Making the most of affinity tags. *Trends in Biotechnology,* 23**,** 316-320.

WEGE, S. & POIRIER, Y. 2014. Expression of the mammalian Xenotropic Polytropic Virus Receptor 1 (XPR1) in tobacco leaves leads to phosphate export. *Febs Letters,* 588**,** 482-489.

WILD, R., GERASIMAITE, R., JUNG, J.-Y., TRUFFAULT, V., PAVLOVIC, I., SCHMIDT, A., SAIARDI, A., JESSEN, H. J., POIRIER, Y., HOTHORN, M. & MAYER, A. 2016. Control of eukaryotic phosphate homeostasis by inositol polyphosphate sensor domains. *Science,* 352**,** 986-990.

WU, F.-H., SHEN, S.-C., LEE, L.-Y., LEE, S.-H., CHAN, M.-T. & LIN, C.-S. 2009. Tape-Arabidopsis Sandwich - a simpler Arabidopsis protoplast isolation method. *Plant Methods,* 5**,** 16.

WU, P., SHOU, H., XU, G. & LIAN, X. 2013. Improvement of phosphorus efficiency in rice on the basis of understanding phosphate signaling and homeostasis. *Current Opinion in Plant Biology,* 16**,** 205-212.

XIANG, Z. X. 2006. Advances in homology protein structure modeling. *Current Protein & Peptide Science,* 7**,** 217-227.

XU, X., PAN, S., CHENG, S., ZHANG, B., MU, D., NI, P., ZHANG, G., YANG, S., LI, R., WANG, J., ORJEDA, G., GUZMAN, F., TORRES, M., LOZANO, R., PONCE, O., MARTINEZ, D., DE LA CRUZ, G., CHAKRABARTI, S. K., PATIL, V. U., SKRYABIN,

K. G., KUZNETSOV, B. B., RAVIN, N. V., KOLGANOVA, T. V., BELETSKY, A. V., MARDANOV, A. V., DI GENOVA, A., BOLSER, D. M., MARTIN, D. M. A., LI, G., YANG, Y., KUANG, H., HU, Q., XIONG, X., BISHOP, G. J., SAGREDO, B., MEJIA, N., ZAGORSKI, W., GROMADKA, R., GAWOR, J., SZCZESNY, P., HUANG, S., ZHANG, Z., LIANG, C., HE, J., LI, Y., HE, Y., XU, J., ZHANG, Y., XIE, B., DU, Y., QU, D., BONIERBALE, M., GHISLAIN, M., DEL ROSARIO HERRERA, M., GIULIANO, G., PIETRELLA, M., PERROTTA, G., FACELLA, P., O'BRIEN, K., FEINGOLD, S. E., BARREIRO, L. E., MASSA, G. A., DIAMBRA, L., WHITTY, B. R., VAILLANCOURT, B., LIN, H., MASSA, A., GEOFFROY, M., LUNDBACK, S., DELLAPENNA, D., BUELL, C. R., SHARMA, S. K., MARSHALL, D. F., WAUGH, R., BRYAN, G. J., DESTEFANIS, M., NAGY, I., MILBOURNE, D., THOMSON, S. J., FIERS, M., JACOBS, J. M. E., NIELSEN, K. L., SONDERKAER, M., IOVENE, M., TORRES, G. A., JIANG, J., VEILLEUX, R. E., BACHEM, C. W. B., DE BOER, J., BORM, T., KLOOSTERMAN, B., VAN ECK, H., DATEMA, E., HEKKERT, B. T. L., GOVERSE, A., VAN HAM, R. C. H. J., VISSER, R. G. F. & POTATO GENOME SEQUENCING, C. 2011. Genome sequence and analysis of the tuber crop potato. *Nature,* 475**,** 189-195.

YANG, S. Y., GRONLUND, M., JAKOBSEN, I., GROTEMEYER, M. S., RENTSCH, D., MIYAO, A., HIROCHIKA, H., KUMAR, C. S., SUNDARESAN, V., SALAMIN, N., CATAUSAN, S., MATTES, N., HEUER, S. & PASZKOWSKI, U. 2012. Nonredundant regulation of rice arbuscular mycorrhizal symbiosis by two members of the phosphate transporter1 gene family. *Plant Cell,* 24**,** 4236-4251.

YUAN, H. & LIU, D. 2008. Signaling components involved in plant responses to phosphate starvation. *Journal of Integrative Plant Biology,* 50**,** 849-859.

YUAN, Z. C., ZAHEER, R., MORTON, R. & FINAN, T. M. 2006. Genome prediction of PhoB regulated promoters in Sinorhizobium meliloti and twelve proteobacteria. *Nucleic Acids Research,* 34**,** 2686-2697.

ZAKHLENIUK, O. V., RAINES, C. A. & LLOYD, J. C. 2001. pho3: a phosphorus-deficient mutant of Arabidopsis thaliana (L.) Heynh. *Planta,* 212**,** 529-534.

ZHANG, K., SONG, Q., WEI, Q., WANG, C. C., ZHANG, L. W., XU, W. Y. & SU, Z. 2016. Down-regulation of OsSPX1 caused semi-male sterility, resulting in reduction of grain yield in rice. *Plant Biotechnology Journal,* 14**,** 1661-1672.

ZHAO, L., LIU, F., XU, W., DI, C., ZHOU, S., XUE, Y., YU, J. & SU, Z. 2009. Increased expression of OsSPX1 enhances cold/subfreezing tolerance in tobacco and Arabidopsis thaliana. *Plant Biotechnology Journal,* 7**,** 550-561.

ZHAO, S. & FERNALD, R. D. 2005. Comprehensive algorithm for quantitative real-time polymerase chain reaction. *Journal of Computational Biology,* 12**,** 1047-1064.

ZHOU, Y. & NI, M. 2010. SHORT HYPOCOTYL UNDER BLUE1 truncations and mutations alter its association with a signaling protein complex in Arabidopsis. *Plant Cell,* 22**,** 703-715.

Appendices

Appendix 1 Codon usage optimized coding sequence of AtSPX1 (by Prof. Stephen Baldwin)

```
  1  M   K   F   G   K   S   L   S   N   Q   I   E   Q   T   L   P   E   W   Q   D
  1  ATGAAATTCGGTAAATCTCTGTCTAACCAGATCGAACAGACCCTGCCGGAATGGCAGGAC
  1  TACTTTAAGCCATTTAGAGACAGATTGGTCTAGCTTGTCTGGGACGGCCTTACCGTCCTG
 21  K   F   L   S   Y   K   E   L   K   K   R   L   K   L   I   G   S   K   T   A
 61  AAATTCCTGTCTTACAAAGAACTGAAAAAACGTCTGAAACTGATCGGTTCTAAAACCGCT
 61  TTTAAGGACAGAATGTTTCTTGACTTTTTTGCAGACTTTGACTAGCCAAGATTTTGGCGA
 41  D   R   P   V   K   R   L   R   L   D   E   F   S   V   G   I   S   K   E   E
121  GACCGTCCGGTTAAACGTCTGCGTCTGGACGAATTTTCTGTTGGTATCTCTAAAGAAGAA
121  CTGGCAGGCCAATTTGCAGACGCAGACCTGCTTAAAAGACAACCATAGAGATTTCTTCTT
 61  I   N   F   I   Q   L   L   E   D   E   L   E   K   F   N   N   F   F   V   E
181  ATCAACTTCATCCAGCTGCTGGAAGACGAACTGGAAAAATTCAACAACTTCTTCGTTGAA
181  TAGTTGAAGTAGGTCGACGACCTTCTGCTTGACCTTTTTAAGTTGTTGAAGAAGCAACTT
 81  K   E   E   E   Y   I   I   R   L   K   E   F   R   D   R   I   A   K   A   K
241  AAAGAAGAAGAATACATCATCCGTCTGAAAGAATTTCGTGACCGTATCGCTAAAGCTAAA
241  TTTCTTCTTCTTATGTAGTAGGCAGACTTTCTTAAAGCACTGGCATAGCGATTTCGATTT
101  D   S   M   E   K   M   I   K   I   R   K   E   I   V   D   F   H   G   E   M
301  GACTCTATGGAAAAAATGATCAAAATCCGTAAAGAAATCGTTGACTTCCACGGTGAAATG
301  CTGAGATACCTTTTTTACTAGTTTTAGGCATTTCTTTAGCAACTGAAGGTGCCACTTTAC
121  V   L   L   E   N   Y   S   A   L   N   Y   T   G   L   V   K   I   L   K   K
361  GTTCTGCTGGAAAACTACTCTGCTCTGAACTACACCGGTCTGGTTAAAATCCTGAAAAAA
361  CAAGACGACCTTTTGATGAGACGAGACTTGATGTGGCCAGACCAATTTTAGGACTTTTTT
141  Y   D   K   R   T   G   D   L   M   R   L   P   F   I   Q   K   V   L   Q   Q
421  TACGACAAACGTACCGGTGACCTGATGCGTCTGCCGTTCATCCAGAAAGTTCTGCAACAG
421  ATGCTGTTTGCATGGCCACTGGACTACGCAGACGGCAAGTAGGTCTTTCAAGACGTTGTC
161  P   F   Y   T   T   D   L   L   F   K   L   V   K   E   S   E   A   M   L   D
481  CCGTTCTACACCACCGACCTGCTGTTCAAACTGGTTAAAGAATCTGAAGCTATGCTGGAC
481  GGCAAGATGTGGTGGCTGGACGACAAGTTTGACCAATTTCTTAGACTTCGATACGACCTG
181  Q   I   F   P   A   N   E   T   E   S   E   I   I   Q   A   E   L   S   E   H
541  CAGATCTTCCCGGCTAACGAAACCGAATCTGAAATCATCCAGGCTGAACTGTCTGAACAC
541  GTCTAGAAGGGCCGATTGCTTTGGCTTAGACTTTAGTAGGTCCGACTTGACAGACTTGTG
201  K   F   M   E   S   L   H   M   K   S   T   I   A   A   L   R   V   L   K   E
601  AAATTCATGGAATCTCTGCACATGAAATCTACCATCGCTGCTCTGCGTGTTCTGAAAGAA
601  TTTAAGTACCTTAGAGACGTGTACTTTAGATGGTAGCGACGAGACGCACAAGACTTTCTT
221  I   R   S   G   S   S   T   V   S   V   F   S   L   P   P   L   Q   L   N   G
661  ATCCGTTCTGGTTCTTCTACCGTTTCTGTTTTCTCTCTGCCGCCGCTGCAACTGAACGGT
661  TAGGCAAGACCAAGAAGATGGCAAAGACAAAAGAGAGACGGCGGCGACGTTGACTTGCCA
241  L   D   E   T   W   K   K   I   P   L   L   E   Q   E   A   K   *
721  CTGGACGAAACCTGGAAAAAAATCCCGCTGCTGGAACAGGAAGCTAAATAA
721  GACCTGCTTTGGACCTTTTTTTAGGGCGACGACCTTGTCCTTCGATTTATT
```

203

Appendix 2 Plasmid maps for *Arabidopsis* SPX domain expression test

pEAQ-HT-His6-AtSPX1
10753 bp

Xma I (1)
Sma I (3)
Misc Feature 2
5'UTR 1
Nco I (354)
Promoter P 1
AtSPX1
Sac II (9900)
Nde I (626)
Misc Feature 1
Xho I (778)
Sph I (9447)
Stu I (784)
Rep Origin 2
3'UTR 1
Pst I (8609)
Pst I (1024)
Terminator 1
Xba I (1278)
Kpn I (1356)
Promoter P 2
CDS 3
Xba I (1769)
Bam HI (1775)
trf A
Cla I (1788)
Nde I (6230)
Kpn I (1869)
Spe I (5948)
CDS 1
Sac II (5809)
Sca I (2531)
Not I (5803)
Terminator 2
Rep Origin 1
Sac II (3149)
Nhe I (3231)
Pst I (3535)
CDS 2
Sph I (3887)
Nco I (3914)
Misc Feature 4

pEAQ-HT-AtSPX1-His6
10750 bp

Nco I (354)
AtSPX1
5'UTR 1
Nde I (626)
Promoter P 1
Xma I (775)
Sac II (9924)
Sma I (777)
Misc Feature 1
Misc Feature 3
Sph I (9471)
Xho I (802)
Rep Origin 2
Stu I (808)
3'UTR 1
Pst I (8633)
Pst I (1048)
Terminator 1
Xba I (1302)
CDS 3
Kpn I (1380)
trf A
Promoter P 2
Nde I (6254)
Xba I (1793)
Spe I (5972)
Bam HI (1799)
Sac II (5833)
Cla I (1812)
Not I (5827)
Kpn I (1893)
Rep Origin 1
CDS 1
Sca I (2555)
Terminator 2
Sac II (3173)
Nhe I (3255)
Pst I (3559)
CDS 2
Sph I (3911)
Nco I (3938)
Misc Feature 4

pGEX-2T-AtSPX1
5715 bp

tac promoter
GST
lacZ a
lac promoter
Sca I (649)
BamHI (748)
Eco RV (4682)
Nco I (1101)
lacI
AtSPX1
HindIII (1261)
BglII (1295)
Nde I (1373)
Eco RI (1525)
AmpR promoter
Sca I (2247)
pBR322 origin
Ampicillin
PstI (2486)

pGEX-2T-AtSPX3
5704 bp

tac promoter
GST
lacZ a
Sca I (649)
Eco RV (4671)
Xba I (1018)
Nco I (1062)
lacI
AtSPX3
Spe I (1224)
BamHI (1260)
Eco RI (1492)
Xma I (1497)
Sma I (1499)
Sal I (1502)
Xho I (1507)
Not I (1513)
AmpR promoter
Sca I (2236)
pBR322 origin
Ampicillin
PstI (2475)

pET28-AtSPX1*-His6
6004 bp

T7 P
Bgl II (5986)
Xba I (48)
Sph I (5797)
Nco I (87)
AtSPX1*
Bgl II (633)
lacI
Xho I (860)
Eco RV (4818)
T7 TERM
f1 ori
Kanr
Xma I (2089)
Sma I (2091)
Cla I (2272)
pBR322 ori

pET28-AtSPX1*-His8
6109 bp

Xba I (1)
Nco I (40)
T7 promoter
AtSPX1*
lac operator
Bgl II (586)
Bgl II (6044)
Pst I (818)
Sph I (5855)
HRV3C cleavage site
Sca I (829)
Nde I (847)
lac I
8His tag
Eco RV (4876)
Bam HI (891)
Eco RI (897)
Hin dIII (903)
Not I (910)
Xho I (918)
His tag
T7 terminator
f 1 origin
kan sequence
Xma I (2147)
Sma I (2149)
ColE1 pBR322 origin
Cla I (2330)

pET28-AtPHO1;H1 1-His6
6256 bp

T7 P
Bgl II (6238)
Xba I (48)
Sph I (6049)
Nco I (87)
Hin dIII (323)
Sac I (528)
lacI
AtPHO1;H1 1
Eco RV (5070)
Pst I (915)
Xho I (1112)
T7 TERM
f 1 ori
Kanr
Xma I (2341)
Sma I (2343)
Cla I (2524)
pBR322 ori

pET28-AtPHO1;H1 2-His6
6388 bp

T7 P
Bgl II (6370)
Xba I (48)
Sph I (6181)
Nco I (87)
Hin dIII (323)
Sac I (528)
lacI
AtPHO1;H1 2
Eco RV (5202)
Pst I (915)
Xho I (1244)
T7 TERM
f 1 ori
Kanr
Xma I (2473)
Sma I (2475)
Cla I (2656)
pBR322 ori

pL53-MBP-AtPHO1;H1 1
6885 bp

-35 trp
-10 lacUV5
lac operator
ptac
Xba I (59)
rbs
MBP signal sequence
RGSH8 tag
Sac II (180)
Bgl II (570)
MBP
Sac II (1159)
Nco I (1311)
Sac I (1327)
Sac I (1366)
Spe I (1393)
Nhe I (1411)
Hin dIII (1645)
Sac I (1850)
AtPHO1;H1 1
Pst I (2237)
Sbf I (2439)
Pst I (2439)
Hin dIII (2449)
rrnB T1 and T2 transcription terminator
pBR322 ori
Ampr
Sca I (5132)
Eco RV (4326)
lacIq

pL53-MBP-AtPHO1;H1 2
7017 bp

-35 trp
-10 lacUV5
lac operator
ptac
Xba I (59)
rbs
MBP signal sequence
RGSH8 tag
Sac II (180)
Bgl II (570)
MBP
Sac II (1159)
Nco I (1311)
Sac I (1327)
Sac I (1366)
Spe I (1393)
Nhe I (1411)
Hin dIII (1645)
Sac I (1850)
AtPHO1;H1 2
Pst I (2237)
Sbf I (2571)
Pst I (2571)
Hin dIII (2581)
rrnB T1 and T2 transcription terminator
pBR322 ori
Ampr
Sca I (5264)
Eco RV (4458)
lacIq

T7 P

Bgl II (6187)

Xba I (48)

Sph I (5998)

Nco I (87)

AtSPX1*

Bgl II (633)

lacI

Xho I (860)

HRV cleavage site

Eco RV (5019)

Nde I (891)

SBP tag

HRV-SBP-HIS8

RGS-His8 Tag

pET28-AtSPX1*-SBP-His8

6205 bp

T7 TERM

f1 ori

Kanr

Xma I (2290)

Sma I (2292)

Cla I (2473)

pBR322 ori

Appendix 3 Plasmid maps for other plasmids used in this research

pL53
6655 bp

-35 trp
-10 lacUV5
lac operator
ptac
Xba I (59)
rbs
MBP signal sequence
RGSH8 tag
Sac II (180)
Bgl II (570)
MBP
MBP-tagged CBZIP
Sac II (1159)
Nco I (1311)
Linker
Sac I (1327)
Sac I (1366)
HRV cleavage site
Linker
Spe I (1393)
Nhe I (1411)
CBZIP
Kpn I (2084)
Afl II (2198)
Pst I (2209)
Hin dIII (2219)
rrnB T1 and T2 transcription terminator
lacIq
Eco RV (4096)
Sca I (4902)
Ampr
pBR322 ori

pMAL-c2x-AtdPHR1
7092 bp

tac promoter
Eco RV (6568)
lacI
Nde I (121)
Bgl II (482)
MalE
Sac I (1233)
Eco RI (1290)
Bam HI (1296)
Eco RV (1524)
AtdPHR1 aa208-362
Xho I (1671)
Hin dIII (1769)
lacZ a
AmpR promoter
Sca I (2765)
Ampicillin
f 1 origin
pBR322 origin
ROP

Appendix 4 Corrected StPHO1;1 and StPHO1;3 amino acid sequences

>StPHO1;1

```
MKFGKEFTSQMVHEWQEAYMDYNYLKNLLKDLLDFKKKNAPLSEVAATPEGSLKRRISMYRAFSGLQSRYTSFKGSSGNNNNNNHEDEVILVNTVQQEG
SQGHHQTMFLMSSEEGGEYEMIFFKRLDDEFNKVSTFYKKKVGQVKAEADELSKQMDALIALRIMVEKPSIEMHSAQGMDPASVVPLHLRSQLRSHMEV
IQEVEMSSEEIVEDESTSGKRDTTKMNPMGFRPAPVGILDNVKINIEPATPISTLKNVIKSSKSDLSFSRQELRKAEEQIRMAFVEFYQKLRLLKNYSF
LNVLAFSKIMKKYDKITSRKASKSYLEMIDKSYLGSSDEVAKLVERVEVTFIKHFVNGNRRKGMKSLRPQAKRDTHRVTFFMGMFSGCSIALVAAIAVV
IRAGNLLEHKDRGQYMDNIFPLYSLFGYIVLHMLMYAGNVYYWKRFRVNYPFIFGFKQGTALGYRQVLLLASGLSLLALAAAFSHLDMDMDPKTRKFET
LTELIPLTLVIVLIIIIFCPLNIIYRSSRFFLLRTAWHCLCAPLYKVTLPDFILADQLTSQVQAIRSLQFYVCYYVWGNFKTRSNKCQDSSVYKILYIV
VAIIPFWSRFIQCLRRLFEEKDSMQGLNGLKYFSTIVALVMKTLYSQKGGTFWRVMAASTSGVTTVANTYWDLVIDWGLLQRNSKNPWLRDKLLVPHKI
VYFVAIVLDIILRLVWMQLVLDIKELSFLHEKAFLAVVACLEILRRGIWNFFRLENEHLNNVGKYRAFKSVPLPFNYDEDKSQ
```

>StPHO1;3

```
MVKFSKQFEGQLVPEWKEAFVDYWQLKKDLKKIHLLNNNVNNANKKSSFSRNIYTSLRKLPLFGPQRRENGIIQVHTKLGQTLSKGDLYETELLEQFAD
TESAAEFFALLDLQLNKVNQFFRTKEKEFIERGECLKKQMEILIELKDALIKQQYDKGTSSGQNIKDDELISATISCDEESNKDRTEQEQDIENSIDQV
ILDSPRSSELGNPTNINTEDNKSKSSSERAINNQGKSLKIHIPLTNPTRTFSAITYLLRDDMINQSSKKCGPNGRQKLHINRTKLKHAEKMIRGAFIEL
YKGLEYLKIYRNLNMLAFVKILKKFDKVTNKQVLPIYLRVVESSYFNSSDKALKLADEVEEIFIKHFAEDDKKKAMKYLKPTQKKESHAVTFFIGLFGG
CFIALLVGYVIMAHITGLYRPKSDTIYMETVYPVLSMFSLMFLHFFLYGCNIFMWRKTRVNYSFIFELAQTKELKYRDVFLICTTSMTAVIGVLFLHLT
LVAKGYSYNQIQAIPALLLLVFILLLVCPFNIIYKSSRYRFIRVIRNIMFSPLYKVVMLDFFMADQLCSQVPMLRNLEYVACYYITGSYKTQDYGYCMR
TKYYRDLAYAVSFLPYYWRAMQCARRWFDEGHKSHLVNLGKYVSAMLAAGAKVAYEKEKNMGWLCLVIVMSSVATVYQLYWDFVKDWGLLQCHSKNPWL
RNELMLRRKFIYYFSMGLNLVLRLAWLQTVLHYNFGTVDYRVTGLFLAALEVIRRGHWNYYRLENEHLNNAGKFRAVKTVPLPFHEVDEQD
```

Appendix 5 Amino acid sequence alignments and phylogenetic trees of potato (St), rice (Os) and *Arabidopsis* (At) SPX-EXS, SPX-MFS and SPX-RING family proteins

The identified potato SPX-EXS, SPX-MFS and SPX-RING family protein sequences were aligned with those already known from rice and *Arabidopsis* using Clustal X (Higgins *et al.*, 1996; Thompson *et al.*, 1997). Members from the SPX-EXS family are only aligned within the 3 sub-domains due to the presence of longer inserted sequences of low similarity in the SPX domains. (Sequences between the sub-domains are represented by dotted lines '--') Conserved amino acid residues from phosphate binding cluster and lysine surface cluster (Wild *et al.*, 2016) are highlighted in yellow and turquoise, respectively. Phylogenetic tree shown overleaf was generated for these homologues using tree view (Page, 1996).

<u>SPX-EXS family members – alignment within the 3 sub-domains</u>

```
                    |————————— sub-domain 1 —————————|
StPHO1;4            -MKFGKEFKKEMVPEWIEAYVDYTGLKQILQEI--
StPHO1;5           -MKFGKKFKDEMVPEWSEAYMDYGGLKHLLKDI--
AtPHO1_H10         -MKFGKIFKKQMVPEWVEAYVDYNGLKRVLKEI--
AtPHO1_H3          -MKFGKEFSSQMVPEWQQAYMDYDFLKTLLKEI--
AtPHO1_H5          -MKFGKEFSSQMVPEWHEAYMDYDYLKSQLKEI--
AtPHO1_H6          -MKFGKDFSS----EWQQAYVDYKYLKTLVKDI--
AtPHO1_H2          -MKFGKELSSQMVQEWQQAYVNYDYLKTLLKEI--
AtPHO1_H7          -MKFGKDFVRQMIPEWQQAYMDYAGLKSILQEI--
AtPHO1_H8          -MKFGKEYVAQMIPEWQQAYMDYTCLKTILREI--
AtPHO1_H4          -MRFGKEFVSQMIPEWQEAYIDYAYLKTILQDI--
StPHO1;1-corrected -MKFGKEFTSQMVHEWQEAYMDYNYLKNLLKDL--
AtPHO1_H9          -MKFGREFETQMIQEWKEAYMDYRSLKSIVKQI--
AtPHO1_H1          MVKFTKQFEGQLVPEWKDASVDYSQLKKDLKKI--
StPHO1;3-corrected MVKFSKQFEGQLVPEWKEAFVDYWQLKKDLKKI--
OsPHO1_3           MVKFSKQFEGQLVPEWKDAFVDYWQLKKDIKRL--
OsPHO1_1           MVKFSKQFEGQLVPEWKHAFVDYSLLKKDLKRM--
StPHO1;2           MVKFSKELEAQLIPEWKDAFVNYWQLKKQVKKI--
AtPHO1             MVKFSKELEAQLIPEWKEAFVNYCLLKKQIKKI--
OsPHO1_2           MVKFSREYEASIIPEWKAAFVDYKRLKKLIKRI--
Clustal Consensus   ::* :        ** * ::* **  :: :
```

```
                                ├———— sub-domain 2 ————┤                    ├————— sub-domain 3 —————┤
StPHO1;4            --FFDKLDHEFNKVNTFYKDKVDEVMREVTLLN--          --KLSHLKHFSYMNLSAFSKILKKYEKIT--
StPHO1;5            --FFEKLDNELEKVNRFYKDIVEEAKGEADELA--          --KLRRLNQYSFMNLSAFSVILEKYEKIT--
AtPHO1_H10         --FFKKLDENLNKVNKFYRDKVKEVIEEAALLD--          --KLRRLKEYSFMNLLAFSKIMKKYEKIA--
AtPHO1_H3          --FFRRLDDEFNKVDKFYRKKVEEVLKEAAMLN--          --KLRLLKSYSFLNVLAFSKILKKYDKIT--
AtPHO1_H5          --FFRRLDDEFNKVEKFYKEKVEEVMKEAVMLE--          --KLRLLKSYSFLNELAFSKILKKYDKIT--
AtPHO1_H6          --FFRRLDDEFNKVEKFYREKVDEVVKEAAVLN--          --KLWYLKSYSFLNVLALSKILTKYDKIT--
AtPHO1_H2          --FFRRLDDEFNRVEKFYKEKVEEVMKDAIMLN--          --KLRLLKSYSFLNVLAFSKILKKYDKIT--
AtPHO1_H7          --FFKTLDLEFDKVNHFYRSKVEEMVKEAVVLN--          --KLRHLKNYSFLNTLAISKIMKKYDKIA--
AtPHO1_H8          --FFKTLDLEFDKVNRFYRSNVEELVKEAVVLN--          --KLRHLKNYSFLNTLAISKIMKKYDKIA--
AtPHO1_H4          --FFRTLDREFNKVNNFYRLKVETARTEALALN--          --KLRHLKNYSFLNASAVSKIMKKYDKIA--
StPHO1;1-corrected --FFKRLDDEFNKVSTFYKKKVGQVKAEADELS--          --KLRLLKNYSFLNVLAFSKIMKKYDKIT--
AtPHO1_H9          --FFRRLDGEFNKVLRFYKQKVENVMEEADELS--          --KLRFLKSYCFLNQLAFAKILKKYDKTT--
AtPHO1_H1          --FFACLDMQLNKVNQFYKTKEKEFLERGECLK--          --GLNYLKTYRNLNILAFMNILKKFDKVT--
StPHO1;3-corrected --FFALLDLQLNKVNQFFRTKEKEFIERGECLK--          --GLEYLKIYRNLNMLAFVKILKKFDKVT--
OsPHO1_3          --FFARLDEQLNKVNRFYERKEAEFVERGESLR--          --GLGYLKTYRSLNMMAFVKILKKFDKVT--
OsPHO1_1          --FFARLDAQLNKVNHFYKAKEEEFLHRGHSLR--          --GLGYLTTYRNLNMMAFVKILKKFEKVS--
StPHO1;2           --FFEMLDEELKKVNEFYKTKESEFLERGDILN--          --GLGLLKTYSSLNMVAFVKILKKFDKVA--
AtPHO1            --FFARLDEELNKVNQFHKPKETEFLERGEILK--          --GLGLLKTYSSLNMIAFTKIMKKFDKVA--
OsPHO1_2          --FMERADEELEKVNAFYTGQEAELLARGDALL--          --GLELLKKFSSLNVKAFTKILKKFVKVS--
Clustal Consensus    *:    * ::.:*  *.          *          * *.: :* *. *: *: * :
```
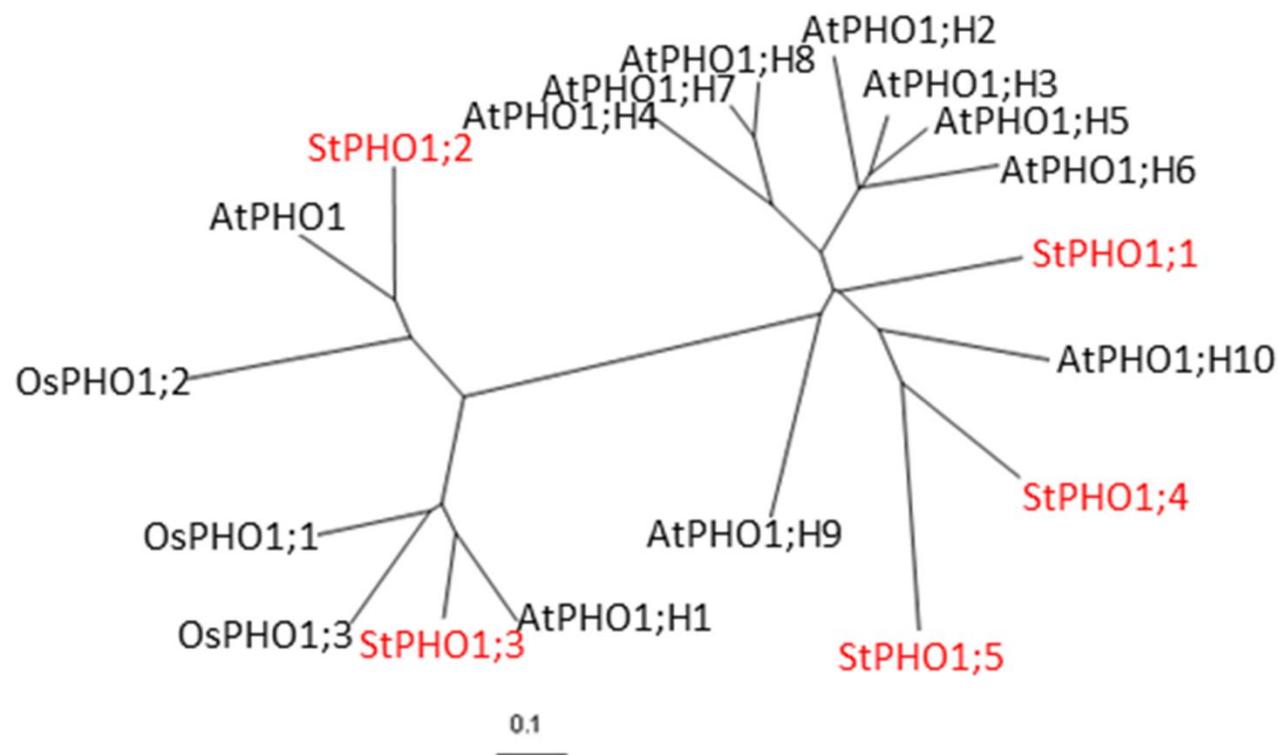
SPX-MFS family members


```
                10        20        30        40        50        60        70
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OsSPX-MFS1  MVNFGKKLMADQVEEWKGYYINYKLMKKMLKQYVQQTQLGGKDREQVLKEFSRILDEQIERIVLFLLQQQ
OsSPX-MFS2  MVNFGKRLMADQLEEWKEYYINYKMMKKKVKQYVQQTQNGGRNREQVLKEFSRMLDDQIEKIVLFLLQQQ
StSPX-MFS1  MVAFGKKLKDRQIQEWQGYYINYKLMKKKVKQYDNQIKAGSLDRRHVLKDFSRMLDNQIERIVLFMLEQQ
StSPX-MFS2  MVSFGKKLKGRQIQEWQGYYINYKLMKRKLKQYANQSQAVVPDRRFVLKDFSRMLDHQIETIVLFLLEQQ
AtSPX-MFS1  MVAFGKKLKERSIQEWQGYYINYKLMKKKVKQYSRQLEGGNLERRHVLKDFSRMLDNQIEKIALFMLEQQ
AtSPX-MFS2  MVAFGKKLKERSIEEWQEYYINYKLMKKKVKQYGPQIEVGSLDRRHVLKDFSRMLDHQIEKIALFMLEQQ
AtSPX-MFS3  MVAFGKYLQRKQIEEWSGYYINYKLMKKKVKQYAEQIQGGSQHPRHVLKDFSRMLDTQIETTVLFMLEQQ
OsSPX-MFS3  MVNFGKKLMADQIPEWKGYYINYKLMKKKVKQYGQQVQQGEKDRRRVLKDFSKMLDDQIEKIVLFLLEQQ
OsSPX-MFS4  MVNFSNKLTKDQIPGWEEYYFNYKMLKGRVNEYTEQTKEGTQYRRRVLKDFSKLLDDEIEKIVLFMIEQQ
Clustal Consensus  ** *.: *    .:  *. **:***::*  ::*  * :     . ***:**::** :**  .**::**

                80        90        100       110       120       130       140
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OsSPX-MFS1  GHLANRIEELGEQRAALLEQHDISQVFQLREAYREVGRDLIKLLRFVDMNATGIRKILKKFDKRFGYRFT
OsSPX-MFS2  GHLASRIEKLGEERALLMEQADASQISELREAYREVGIDLMKLLRFVDMNATGIRKILKKFDKRFGYKFT
StSPX-MFS1  GVLASRISELNKQQESLQEQPDISKIIELRESYRDVGRDLLKLLFFVEINAIGLRKILKKFDKRFGYRFT
StSPX-MFS2  GALASRISELNEQKDSLQEVPDISKIDELREVYQAVGRDLLKLLYFVEINAIGLRKILKKFDKRFGYKFT
AtSPX-MFS1  GLLASRLQTLRGSHDALQEQPDISHMSYLKEEYRAVGQDLLKLLFFVEMNAIGIRKILKKFDKRFGYRFT
AtSPX-MFS2  GLLSSRLQKLREWHDTLQDEPDLSQIAKLREAYRAVGQDLLKLLFFIDMNAIGIRKILKKFDKRFGYRFT
AtSPX-MFS3  GLLSGRLAKLRESHDAILEQPDISRIFELREAYRDVGRDLLQLLKFVELNAIGLRKILKKFDKRFGYRFA
OsSPX-MFS3  GALASRIEKLGKQRAILAEQPDISAIAELREAYREVGLDLIKLLKFVDLNATGIRKILKKFDKRFGYRFT
OsSPX-MFS4  GLIAARLEDLGKRRARLQDIPLLQEITELREDYRSVGLDLVTLLKFVELNANAVRKILKKFDERLGYKFT
Clustal Consensus  * :: *:  *    :  : :     . :  *:* *: ** **: ** *::**  .:*******:*:**:*:

                150       160       170       180       190       200       210
           ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OsSPX-MFS1  DYYVTTRANHPYSQLQQVFKQVGIVAVVGALSRNLAYLQDHEGSVLSIYDHPSVTL-KDPIIDQVNHAVQ
OsSPX-MFS2  DYYVSTRANHPCSQLQQIFKQVGIVAVVGALSRNLAFLQDHQGNFPSIYDHPSITL-KDPIIEQINHSVQ
StSPX-MFS1  DYYVKTRANHPYSQLQQVFKHVGLGAVVGAISRNLADLQDRQGSYLSIYDQPSLPL-QDSVVDSLQAAVD
StSPX-MFS2  DYYVKTRANHPYSQLQQVFKHVGLGAVVGAISRNLADLQDRQGSYLSIYDQPALPL-QDPVVDSMQAAVD
AtSPX-MFS1  NYYVKTRANHPYSELQQVFRHVGLGAVVGAVSRNLHELQNNQGSYLSIYDQPVLPL-QDPVVDSIRAAVD
AtSPX-MFS2  NYYVKTRADHPYSQLQQVFRHVGLGAVVGAISRNLHELQNNEGSYLSIYDQPVLPL-QDPVVDSIKNAVD
AtSPX-MFS3  DYYVKTRANHPYSQLQQVFKHVGVGAVVGAISRNLHELQENEGSFYSIYDQPVLPA-QDPVVEAINNAVD
```

```
OsSPX-MFS3        DYYVTSRSNHPYSQLQQVFKHVGVGAVVGALSRNLADLQERQGSYLSIYDQPSTAL-KDPIIDMINSSVD
OsSPX-MFS4        DYYVRSRSNHPYSQLQQVFRHVGIGAVVGALSRNLSDLEERQGSYLNIYDQHPLAIPKDPIIDLITATAD
Clustal Consensus :*** :*::** *:***:*::**: *****:****  *::.:*.  .***:   .  :*.::: :  :.:
```

```
                            220       230       240       250       260       270       280
                   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OsSPX-MFS1        KLTHATSFLQFLGQHALIIQED---VQSGSEDLVDDQ--SYHFMSLILNLVNTFLYMVNTYIIVPTADDY
OsSPX-MFS2        KLTHATNLLQFIGQHALIIPED---MHSGSEDLVDDQ--SYHFMSLLLNLANTFLYMVNTYIIVPTADDY
StSPX-MFS1        RLSHSTNFLNFLAQHALIMQEE---FPTPVEEQVDDQ--RYHFMSLILNLANTFLYMVNTYIIVPTADDY
StSPX-MFS2        RLSHSTNFLNFLAQHALIMQEE---LPAPVEEGVDDQ--RYHFMSLLLNLANTFLYMVNTYIIVPTADDY
AtSPX-MFS1        RLTRSTNFLHFMAQHALIMQEE---LPSPQDEEGEEDGRYHFMSLLLNLVNTFLYMVNTYIIVPTADDY
AtSPX-MFS2        RLTHSTNFLNFMAQHALIMQDDEDLLMLPPDEQAEKEEGRYHFMSLLLNLANTFLYMVNTYIIVPTADDY
AtSPX-MFS3        KLTFSTNFLNFLAQHALIMQDD---LVTPSEDTIDER--SYHFNSLLLNLGNTFLYMVNTYIIVPTADDY
OsSPX-MFS3        KLTRSTNFLRFLGQHALIVGEES--PSTAEEEEIEDQ--KYHFMSLMLNLVNTFLYMVNTYIIVPTADDY
OsSPX-MFS4        KLTNSTNFLRFLGQHALIAQADS-TAGTEDEQHVGED--EYHLMSLVLNLANTFLYMVNTYIVVPTADGY
Clustal Consensus :*: :*.:*.*:.*****   :         ::   .    **: **:*** **********:*****.*
```

```
                            290       300       310       320       330       340       350
                   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OsSPX-MFS1        AVSLGAAATVCGVIIGSMAVAQVFSSVYFSAWSNRSYFRPLVFSSIMLFAGNLLYALAYDLNSLTVLLIG
OsSPX-MFS2        SVSLGAAATVCGVIIGSMAVAQVFSSVYFSAWSNKSYFRPLVFSSIMLFLGNLLYALAYDVNSLTVLIVG
StSPX-MFS1        SMSLGAAATVCGIVIGAMAVAQIFSSVYFSAWSNRSYFRPLVFSSIVLFVGNIMYALAYDLKSIPVLLIG
StSPX-MFS2        SMSLGAAATVCGIVIGAMAVAQIFSSVYFSAWSNRSYFRPLIFSSIVLFIGNVMYALAYDLNSIPVLLIG
AtSPX-MFS1        SMSLGAAATVCGVVIGAMAVAQLFSSVYFSAWSNRSYFKPLIFSSIVLFIGNLLYALAFDFNSIAVLLIG
AtSPX-MFS2        SMSLGAAATVCGVVIGAMAVAQLFSSVYFSAWSNKSYFKPLIFSSIVLFFGNLLYALAYDFNSLALLLIG
AtSPX-MFS3        SMSLGAAATVCGVVIGSMAVAQVFSSVYFSAWSNKSYFKPLVFSSIALFIGNLMYALAYDANSIALLLLG
OsSPX-MFS3        SVSLGAASTVCGVVIGSMAVAQIFSSVYFSAWSNKSYFRPLIFSSIVLFLGNVCYAMAYDMKSLTVLIIG
OsSPX-MFS4        ATSLGAAATACGAVIGSMAVAQVFSSVYFSAWSNRSYFRPLLFSSVVLLLGNVMYAMAFDLGSLTILLLG
Clustal Consensus :  *****:*.** :**:*****:*********:***:**:*** *: **: **:*:* *:..*::*::*
```

```
                            360       370       380       390       400       410       420
                   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OsSPX-MFS1        RLLCGLGSARAVNRRYISDCVPLKIRLQASAGFVSASALGMACGPALAGLLQTRFKIYSLTFDQSTLPGW
OsSPX-MFS2        RLLCGLGSARAVNRRYISDCVPLKTRLQASAGFVSASALGMACGPALAGLLQTNFKIYGFTFDQNTLPGW
StSPX-MFS1        RIFCGLGSARAVNRRYISDCVPLKIRMQASAGFVSASALGMACGPALAGLLQTNFKIYKLTFNQDTLPGW
StSPX-MFS2        RLFCGFGSARAVNRRYISDCVPLKYRMQASAGFVSASALGMACGPAVAGLLQTNFKIYKITVNKETLPGW
AtSPX-MFS1        RLFCGLGSARAVNRRYISDCVPLKIRMQASAGFVSASALGMACGPALAGLLQIRFKIYKLTFNQDTLPGW
AtSPX-MFS2        RLFCGFGSARAVNRRYISDCVPLKIRMQASAGFVSASALGMACGPALAGLLQTDFKIKNVTFNQDTLPGW
```

```
AtSPX-MFS3       RVCCGLGSARAVNRRYISDCVPLRIRMQASAGFVSASALGMACGPALAGLLQIKFKFYKFTFNQSTLPGW
OsSPX-MFS3       RLLCGMGSARAVNRRYISDCVPARIRMQASAGFVSASALGMACGPALAGLLQWKFKIYMVTFNQSTLPGW
OsSPX-MFS4       RVLCGMGSARAVNRRYISDCVPPRIRMQASAAFVSASALGMACGPALAGLLQTNFSLYGLTINQITLPGW
Clustal Consensus *: **:************** : *:****.**************:*****  *.:  .*.:: *****
```

```
                          430       440       450       460       470       480       490
                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OsSPX-MFS1       VMCIAWLVYLLWLWISFKEPGHFAKSSDTAQP-AESG-------HQVNANLEEGLAQPLLTGSEEGQDQ-
OsSPX-MFS2       IMCLAWITYLFWLWISFKEPDHIVRENSVNTPSSDSG-------HRRNSNLEDGLAQPFLIDAKESLDE-
StSPX-MFS1       LMAFAWLVYLIWLWISFREPATETE--VNTGPQEPA-------NNVENDVLENGVTQPLLLKVEDNQQDD
StSPX-MFS2       LMAIAWLIYLVWLWFSFKEPVRDTE--IKNVPQE---------SKAEPDSLEKGIVEPLLLKSPENQQD-
AtSPX-MFS1       VMAVAWLIYLVWLAISFREPAREPE--EIPKTSEES-----NHSAVQDVNLEKGMKQPLLLTSEEIEEQ-
AtSPX-MFS2       VMAVAWLLYLVWLAISFREPAREPE--EIHTSQESTSEQIFCGEADQDGNIEKGLKKPLLLASEETEHD-
AtSPX-MFS3       VMAVAWLFYLVWLCISFREPLRDTEDGEKNNRNETTS------DRVESSRVEEGLRLPLLITSGIKPED-
OsSPX-MFS3       VMAVAWLLYLVWLWISFKEPNRATEVNGTQQNPASVQ-------RADIEQLENGLAQPLLRDSSKKDED-
OsSPX-MFS4       IMAFGWLVYLIWLWILFQEPDLGPDVKDFYEGSSSSTS----TRYMEQEKMEQGFTEHLLPSEQDEEDD-
Clustal Consensus :*...*: **.** : *:**                            :*.*.   :*        .:
```

```
                          500       510       520       530       540       550       560
                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OsSPX-MFS1       -NAEDNDDNEEESKNSHGPATSISSAYKLLTPSVKVQLLIYFMLKYAMEILLSESSVITTYYFNWSTSAV
OsSPX-MFS2       -NGEDNDENEEDPEDSHKPATSLAAAYRLLTPSVKVQLLIYFMLKFAMEILLSESSVVTTFYFNWSTSTV
StSPX-MFS1       ENDQEYDESEEAPEESRQPANSIAEAYRLLTPSVKVQLLIYFMLKYAMEVLLSESSVVTTYYFSWSTGTV
StSPX-MFS2       EDEQEGDESEEAPEDSHKPANSIVAAYRLLTPSVKVQLLIYFMLKYAMEILLSESSVITTYYFSWSTGHV
AtSPX-MFS1       -GEDECDGSEEASEDSRTPANSILAAYRLLTPSVKVQLLIYFMLKYAMEILLSESSVITTYYFGWSTSSV
AtSPX-MFS2       -EEDDGDGSEESSDDSRKPANSFVAAYKLLTPSVKVQLLIYFMLKYAMEILLSESSVVTTYYFGWSMSSV
AtSPX-MFS3       --EEECDESEESPEDSHKPANSFIEAYRLLTPSVKVQLLIYFMLKYSMEILLSESSVITSYYFSWTTSSV
OsSPX-MFS3       -DDEEVDDSEEGTHDSRKPATSIGSAYRLLTPSVKVQLLIYFMLKYAMEILLSESSVITNHYFNWNTSAV
OsSPX-MFS4       NGDEEHNETLSSSTTTLRPASSVASAYTLLTPSVKVQLLIYFMLKYAMEILLAESSVVTGYYFGWDIGTV
Clustal Consensus   :: : . . .  :  **.*.  ** *****************:**:**:****:* .**.*  . *
```

```
                          570       580       590       600       610       620       630
                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OsSPX-MFS1       AIFLAILGCTVLPVNAIVGSYITNLFEDRQILVASEIMVLIGIIMSFRYTPHYSVPQYVLSALITFVFAE
OsSPX-MFS2       AMFLAVLGLTVLPVNVIVGSYVTNLFQDRQILVASEIMVLIGIAMSFRFTSHYSVPQYVSSALITFVFAE
StSPX-MFS1       AIFLACLGLTVLPINVVVGSYISNMFEDRQILLVSEIMVFLGILGSFHIVSPYTVPQYVCSGLIMFVAAE
StSPX-MFS2       AVFLACLGLTVLPVNLVVGSYITNMFEDRQILLASEIMVCVGILMSFNVIIPYSVPQYVCSGLLLFVSAE
AtSPX-MFS1       AIFLFCLGLTVLPVNLVVGSYISNMFEDRQILLVSEIMVCVGILLSFHVVVPYTVPQYVCSGLIMFVSAE
```

```
AtSPX-MFS2       SIFLFCLGLTVLPVNLVVGSYISNMFEDRQILLASEIMVCIGIVLSFHVVIPYTVPQYVISGFIMFVSAE
AtSPX-MFS3       AIFLACLGLTVLPINILVGSYISNMFEDRQILLTSEIIVFLGILFSFNLFPYTVPQYVISGLIMFVAAE
OsSPX-MFS3       AIFLAILGLTVLPVNAVVGTYISNMFEDRQLLMVSQITLLVGIIFSFKITSTYSVVQYVVSALVTFVSAE
OsSPX-MFS4       SVFLAVLGLSVLPVNAIVGTYISNMFEDRQILVASEMALLAGVMLSFKLTVEYTAAQYVCSAVLTFVSAE
Clustal Consensus ::**   ** :***:* :**:*::*:*:***:*:.*:: :  *:  **.    *:. *** *..: ** **
```

```
                          640       650       660       670       680       690       700
                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
OsSPX-MFS1       VLEGVNLSLLSRVMSSRLARGTYNGGLLSTEAGTLARVVADATITAAGYLG-PDLLLNITLLPPLVICIA
OsSPX-MFS2       VLEGVNLSLLSRVMSSRLSRGTYNGGLLSTEAGTLARVAADMTITAAGYLG-QNSLLNVTLLPSFVICVA
StSPX-MFS1       VLEGVNLSLLSRVMSSRLSRGTYNGGLLSTEAGTFARVIADVTITLAGYLG-QSKLLNVTLLPSLLICLA
StSPX-MFS2       VLEGVNLSLLSRVMSSRLSRGTYNGGLLSTEAGTIARVIADATITLAGYLG-ESMLLNITLLPSLFICIV
AtSPX-MFS1       VLEGVNLSLLSRVMSSRLSRGTYNGGLLSTEAGTIARVIADATITVAGFFG-RNMLLNVTLLPSLVICVL
AtSPX-MFS2       VLEGVNLSLLSRVMSSRLSRGTYNGGLLSTEAGTIARVIADATITLAGFLG-QSMLLNVTLLPSLIICVL
AtSPX-MFS3       VLEGVNLSLLSRVMSSRLSKGTYNGGLLSTEAGTLARVVADATITLGGYLG-RGHLLNATLLPSLVICIG
OsSPX-MFS3       VLEGVNLSLLSSVMSSRLSRGTYNGGLLSTEAGTLARVVADCTITAAGYLG-IGKLLNVTLLPSLVICAA
OsSPX-MFS4       VVEGVNLSLLSRVMSARLSRGTYNGGLLSTEAGTVARVVADGTITAAGLLAGEGRLLNATLLPALLVCVA
Clustal Consensus *:********* ***:**::**************.*** ** *** .* :.   . *** ****.:.:*
```

```
                          710
                 ....|....|....
OsSPX-MFS1       SLVATFCTYNTLY-
OsSPX-MFS2       SIVATFCTYNSLY-
StSPX-MFS1       SIIATCYTYNSLY-
StSPX-MFS2       SILATFWTYNSLY-
AtSPX-MFS1       SIVATCFTYNSLY-
AtSPX-MFS2       SILATCYTYNSLY-
AtSPX-MFS3       SIVATCCTYNSLY-
OsSPX-MFS3       SIASTFLTYNSLF-
OsSPX-MFS4       SIAATLSTYNSLFY
Clustal Consensus *: :*  ***:*:
```

SPX-RING family members

```
                    10        20        30        40        50        60        70
               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
StNLA1         MKFGETFMEYLQA-EEG--LDKFSPHVEYKRLKKVLKSCRACRAAALKESNSNGEQQEDHENEGSEICQL
AtNLA2         MKFGETFTEYLHG-EEEWFLEKCR-FVEYKKLKKVLKKCKTCNS---TKSDDGQIIPSATSSSLSDSCEC
OsNLA2         MKFGAIYEEYLRE-QQDKYLTKCS-HVEYKRLKKVLKKCRVGRS--LQEDCPNGDQQEGN-NESPDICKC
OsNLA1         MKFAKKYEKYMKG--MDEELPGVG----LKRLKKLLKKCRSDLQ----SHENDGSSAG--------RCPG
StNLA2         MKFCKKYEEYMEGQCQNKKLPRVG----LKKLKKILKKCRKCHQ----SRSVVGLSADNNNVHDSSSCSQ
AtNLA          MKFCKKYEEYMQGQKEKKNLPGVG----FKKLKKILKRCRRNHV----PSRISFTDAIN------HNCSR
Clustal Consensus  ***   : :*:.        *        *:***:** *:                         *


                    80        90        100       110       120       130       140
               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
StNLA1         ESCQLCDQNFFAELKKEASDIAGCFSSRVRRLLQLHTAPGIQKYLVTLR-QCFKNDQQAMMQECQILIEY
AtNLA2         KACPWCDQMFFEELMKEATDIAGFFRSRVRHLLHLHVATGMQRYMIRLR-RCFTDEKQALVQEGQILIQY
OsNLA2         NSCTLCDQMFFTELTKEASEIAGCFSSRVQRLLNLHVPSGFLRYIWRVR-QCFIDDQQIMVQEGRMLLNY
OsNLA1         H-CSVCDGSFFPSLLNEMSAVIGCFNEKAKKLLELHLASGFKKYTMWFTSKGHKS-HGALIQQGKDLVTY
StNLA2         HQCSVCDGSFFPSLLKEMSEVVGSFNKRAQKLLELHLSSGFRKYLILLKEKIQGN-HIALVQEGKDLVTY
AtNLA          E-CPVCDGTFFPELLKEMEDVVGWFNEHAQKLLELHLASGFTKCLTWLRGNSRKKDHHGLIQEGKDLVNY
Clustal Consensus  . *   **    **  .* :*     :  * *  .:.::**.** ..*: :     .   .  . :   ::*: : *: *


                    150       160       170       180       190       200       210
               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
StNLA1         AMMNAIAMQKILKKYDKVHCSVTGRNFKSKMRSERLEILQSPWLIELGALYMNFNES----NGGKS----
AtNLA2         ITMNAIAIRKILKKYDKVHSSENGKNFKLKMRAERIELLHSPWLIELGAFYLNSGLD----NVGNF----
OsNLA2         VTMNAIAIRKILKKYDKIHGSVSGRDFKSKMQTDHIELLQSPWLIELGAFHLNCNSSDIDETVGFL----
OsNLA1         AIINAVAMRKILKKYDKIHYSKQGQEFKAQAQSLHIEILQSPWLCELMAFYMNLRRS--KKNNG------
StNLA2         AIINAIAVRKILKKYDKIHYSKQGQAFKSQAQSMHVEILQSPWLCELMAFHINLREN--KAKNGK-----
AtNLA          ALINAVAIRKILKKYDKIHESRQGQAFKTQVQKMRIEILQSPWLCELMAFHINLKES--KKESGATITSP
Clustal Consensus   :**:*::*********:* *   *: ** : :   ::*:*:**** ** *:::*    .       *


                    220       230       240       250       260       270       280
               ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
StNLA1         ---NVIFSQFFCN-LSDTGSIMTLKFPDSVKLEYDLTCPICLDTVFNPYALSCGHLFCKSCACTAASVMI
AtNLA2         ---KNSFGRVACENLNEDQPVLKLMLPNSIELEYDLTCAICLETVFNPYALKCGHIFCNSCACSAASVLI
OsNLA2         ---KNEFFKNFSCDLTEARPLMTMAISETMKYEYSLTCPICLDTLFNPYALSCGHLFCKGCACGAASVYI
```

221

```
OsNLA1           ---AMELFGDCSLVFDDDKPTISCNLFDSMRVDISLTCSICLDTVFDPVALSCGHIYCYLCSCSAASVTI
StNLA2           ---ALALFDGCLLVFNEGKPSLTCELFDSIKLDIDLTCSICLDTVFDPVSLTCGHIFCYICACKVASVTI
AtNLA            PPPVHALFDGCALTFDDGKPLLSCELSDSVKVDIDLTCSICLDTVFDPISLTCGHIYCYMCACSAASVNV
Clustal Consensus    :       : :   . :.   : :::. :  .***.***:*:*:* :*.***::*  *:* .*** :


                       290       300       310       320       330       340       350
                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
StNLA1           FQGIKAASNVSKCPVCREVGTYANAVHMLELDLLQKKRFKQYWKERHASERAEMVKQSKIYWDNQTRYAV
AtNLA2           FQGIKAAPRHSKCPICREAGVYAEAVHMIELHLLLKTRSKEYWKERMMNERSEMVKQSKMFWNEQTKHMI
OsNLA2           FQGVKSAPPEAKCPVCRSDGVFAHAVHMTELDLLIKTRSKDYWRQRLREERNEMVKQSKEYWDSQAMLSM
OsNLA1           VDGLKSAERKSKCPLCRQAGVFPNAVHLDELNMLLSYSCPEYWEKRIQMERVERVRLAKEHWESQCRAFL
StNLA2           VDGLQAANHKEKCPLCREKGVYESAVHLEELNILLSRSCPEYWERLQTERVERLRLAKEHWESQCRAFM
AtNLA            VDGLKTAEATEKCPLCREDGVYKGAVHLDELNILLKRSCRDYWEERRKTERAERLQQAKEYWDYQCRSFT
Clustal Consensus .:*:::*     ***:**. *.:   ***: **.:* .     :**.:*    ** * :: :* .*: *


                 ..
StNLA1           GF
AtNLA2           GY
OsNLA2           GI
OsNLA1           GM
StNLA2           GV
AtNLA            GI
Clustal Consensus *
```

222

Appendix 6 Sequence alignment of the all AtPHO1 family members (originally done by Prof. Stephen Baldwin)

Totally 11 Arabidopsis PHO1 family protein sequences were aligned using Clustal X (Higgins *et al.*, 1996; Thompson *et al.*, 1997). Sequences for AtPHO1;H1_1 and AtPHO1;H1_2 are highlighted in yellow and underlined, respectively.

```
                       |---------------- conserved SPX domain ----------------|
                            10        20        30        40        50        60        70
                       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AtPHO1           MVKFSKELEAQLIPEWKEAFVNYCLLKKQIKKIKTSRKPKPASH--------------------------
AtPHO1_H1        MVKFTKQFEGQLVPEWKDASVDYSQLKKDLKKIHLFTNGVEKKH--------------------------
AtPHO1_H2        -MKFGKELSSQMVQEWQQAYVNYDYLKTLLKEIIKLKEKTN----PPPPPHHAVPGEG------------
AtPHO1_H3        -MKFGKEFSSQMVPEWQQAYMDYDFLKTLLKEIITFKRRTN-----NAPSHGGAKTGGG-----------
AtPHO1_H4        -MRFGKEFVSQMIPEWQEAYIDYAYLKTILQDIQASRNRSDS----------------------------
AtPHO1_H5        -MKFGKEFSSQMVPEWHEAYMDYDYLKSQLKEIIKFKRKTN----PHGPHHHHHHL--------------
AtPHO1_H6        -MKFGKDFSS----EWQQAYVDYKYLKTLVKDINRFKRKTN-----------------------------
AtPHO1_H7        -MKFGKDFVRQMIPEWQQAYMDYAGLKSILQEIQTSRKRSERP---------GI----------------
AtPHO1_H8        -MKFGKEYVAQMIPEWQQAYMDYTCLKTILREIKTSQKRSESQ---------GV----------------
AtPHO1_H9        -MKFGREFETQMIQEWKEAYMDYRSLKSIVKQILRYRLQKQQRPPPPPPPPSTGDTVPLKTDGGEGGGGG
AtPHO1_H10       -MKFGKIFKKQMVPEWVEAYVDYNGLKRVLKEIRSYKHSKLTR------AASRV----------------
Clustal Consensus    ::* :        ** :* ::*  **   ::.*
```

```
                       |---------------- conserved SPX domain ----------------|
                            80        90        100       110       120       130       140
                       ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AtPHO1           ----YPIGHHSDFGRSLFDPVR---------KLARTFS-------DKLFSNS--EKPEILQVRRRRGSSE
AtPHO1_H1        ------------TETSLIKTVK--------SSLGR----------LSIFGNKGREQSRVIQVHKKLASSG
AtPHO1_H2        ----------ISRKMTLYRAFSGLVQTPGKKRQSSRQ------------------SNYSSEIDIEEGKAP
AtPHO1_H3        ----------LNRKLTLYRAFSGLVSTPRHKRSNSSH-----------------DVEEGVQLTGSMRSGP
AtPHO1_H4        -----------NNQSSTPSFARNLTRRYNRDALVS------------------ENHDIVVNTVTRL
AtPHO1_H5        ----------LHRKMTLHRAFSGLISTSPKKKKHHGHGGGHGGGH-QIGHFS--DSDDDIEEGIKHVTAP
AtPHO1_H6        ----------------LHGGQISLSST---------------------------VLEIEDGITTAT
AtPHO1_H7        ----------LKRKLSGSRNFSG-LTKRYSRTASTRE----------------PEIQDILVHATTG-
AtPHO1_H8        ----------LKRKLSGRRNFSG-LTKRYSRTFSSRD----------------LENHDIMVHATTG-
AtPHO1_H9        GGGG----PGLSRRISLYRAFSGLTNRASRSPKKSHKHHNP-----LSSKRHHHHHNHHHYHLFDDDEEQ
AtPHO1_H10       ----------SQQAEALHRSFSGLSFHPRHSERAG--------------------DIEDQVIKVDTVQ
```

Clustal Consensus

────────────────────── conserved SPX domain ──────────────────────
```
              150       160       170       180       190       200       210
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AtPHO1     TGDDVDEIYQTE--LVQLF--SEE--DEVKVFFARLDEELNKVNQFHKPKETEFLERGEILKKQLETLAE
AtPHO1_H1  S---NNDVYETE--LLEK---IADDTDAAKEFFACLDMQLNKVNQFYKTKEKEFLERGECLKKQMDILIE
AtPHO1_H2  IL---VS-KSTHGLETTFLMTAEEGGEYELVFFRRLDDEFNRVEKFYKEKVEEVMKDAIMLNKQMDALIA
AtPHO1_H3  IL---VNTTASHGYETTFLMAAEEGGEYELVFFRRLDDEFNKVDKFYRKKVEEVLKEAAMLNKQMDALIA
AtPHO1_H4  -----EEGLETAAYETTFLKAGEAGGDFEVTFFRTLDREFNKVNNFYRLKVETARTEALALNKQMDALIA
AtPHO1_H5  IL---IN-SASHGYETTFLMASEEGGEYETVFFRRLDDEFNKVEKFYKEKVEEVMKEAVMLEKQMDALIA
AtPHO1_H6  IQ---VSSTASQRYETTFLMTAEKGGEYELVFFRRLDDEFNKVEKFYREKVDVVKEAAVLNKQMDALIA
AtPHO1_H7  -----DDGFER--YETTILEVAEAGRESELAFFKTLDLEFDKVNHFYRSKVEEMVKEAVVLNKQMDALIA
AtPHO1_H8  -----DDGFEK--YETTILKVSEVGRESELVFFKTLDLEFDKVNRFYRSNVEELVKEAVVLNRQMDALIA
AtPHO1_H9  IIL--INEDETASYTTTFLNSAEEGGEMEVQFFRRLDGEFNKVLRFYKQKVENVMEEADELSRQLNVLIA
AtPHO1_H10 ------EEGSRKLYETKFLKKSEEGGEFEESFFKKLDENLNKVNKFYRDKVKEVIEEAALLDKQMDALIA
Clustal Consensus   .              :      **  ** ::::* .*:: :       .  *.:*:: *
```

────────────────────── conserved SPX domain ──────────────────────
```
              220       230       240       250       260       270       280
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AtPHO1     LKQILSD----------RKKRNLSGSNS-HRSFS---SSVRNSD---FSAGSPG-----ELSEIQSET-S
AtPHO1_H1  LKDAFKQKQ---ANG--ESTQESKEDDSISCTISCEYDSVRG-----RTEEMQL-----QVSCLDNLE-D
AtPHO1_H2  FRVKVENPVGWG---WEERTVEMTRLAS-DIATSTAAIAASTPARTR---------TMNPRAQAHMEAIQ
AtPHO1_H3  FRVKVENPDGWR---WEERTVEMTRLAS-DIATSAAALSASTPAGAK---------SMKVRSQEHMEAIQ
AtPHO1_H4  FRHKVMD--------QNQKNPSVFDSVS-EDINGSASEVGS--------------------SSKCT
AtPHO1_H5  FRVKVEHPDGWP---WEERTVEMTQLAS-DVANSAAAVAASTPAGAR---------SMKVGAQAHMEAIQ
AtPHO1_H6  FRLKMK---------EESTVEMARFAL-HGVVSPAELAKNP-----------SMKV----HMEAIE
AtPHO1_H7  FRIKVE--------RPSSSWSCSETVS-VDMNALDSNDQR--------------------NTLAE
AtPHO1_H8  YRIKLD---------QPSTSWSCSETVS-VDINALDSKEQKG------------------KTLAE
AtPHO1_H9  LKVKVE-----------NPHVHLPPDLN-SVASAPSSPHSTMR------------------TPAPS
AtPHO1_H10 LRVKMQKP--------DVDNLNLEKHPS-DKVVVDTSDNTMRTQGTA-------NTDMVHGIERTNIPEE
Clustal Consensus  :  .
```

────────────────────── conserved SPX domain ──────────────────────
```
              290       300       310       320       330       340       350
         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AtPHO1     RTDEIIEALERNGVSFINSA--------------------TRSKTK-GGKPKMSLRVDIPDAVAGAEGG
```

```
AtPHO1_H1    NGEEALESLGSEEPIKANNEDSK----------------LTTVSSRVFSCQGK-NVKIKIPLTNP---SR
AtPHO1_H2    EGSFSRE---NEDEDHGSVRGATGDVKTSSLN------TMRGARPAPIEVLDH----IKINNTKA----T
AtPHO1_H3    EGGSSRAGLMEDDEEDEDEQNETSVVSTGAIDNETTTSRMRGARPSPIDVLGR----VKINNTKE----T
AtPHO1_H4    EHNVALADLMRNE-DTSNE----------------------------SILER----IRMNKTREI---T
AtPHO1_H5    EGGSSKAGKSSDEEDDDDAEKEEDNGVSGEVSGDVR--KMKAGRPPPIEVLDR----VKFNHTKE----T
AtPHO1_H6    EGGSSRAGRRSD-EDDYYTDEEDHNDVFFTPANNLS--KMKSSSSAFIEVLDS----IKINNTKE----A
AtPHO1_H7    EMGIRVEGNGSNGGDSTKES--------------------VPQVLSVLER----IRLNKTQE----T
AtPHO1_H8    EMGIKVEENVSNGGDSTKET--------------------APEALSVLDR----IRLNKNQE----N
AtPHO1_H9    PMDVIREMEKTEDKKV---------------------FKPAPVEMLDH----VKLKIDPE----T
AtPHO1_H10   EASHIMADIVPVSHTNGDEEEASIG--------------DKQDLREILER----VKMNDVLE----S
Clustal Consensus                                                           :  :
```

───────── conserved SPX domain ─────────

```
                360       370       380       390       400       410       420
             ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AtPHO1       IARSIATAMS-VLWEELVNNPRSD----FTNWKNIQSAEKKIRSAFVELYRGLGLLKTYSSLNMIAFTKI
AtPHO1_H1    TFSAISYLIN--QSSSKKNGPDG-GNKLQISKKKLSHAEKMIKGALTELFKGLNYLKTYRNLNILAFMNI
AtPHO1_H2    PRSTIKGVLNSSSQNEIIFNRQN-----------LNEVEEKLKFAFVEFYQKLRLLKSYSFLNVLAFSKI
AtPHO1_H3    PRSTIKGVLKVSKQTDLKFSREN-----------LMKVEESLKRAFIEFYQKLRLLKSYSFLNVLAFSKI
AtPHO1_H4    PLSAIKTILKVHKQDELKFTRDN-----------LKEVEKRLQVAFIEFYQKLRHLKNYSFLNASAVSKI
AtPHO1_H5    PRSTIKSVLQASNLTELKFSREN-----------LRKVEAKLRRAFVEFYQKLRLLKSYSFLNELAFSKI
AtPHO1_H6    LQSNTKSVLKVSNHTELKFSRDN-----------LRKIEEKLICAFVEFHRKLWYLKSYSFLNVLALSKI
AtPHO1_H7    PLSTIKNVLKLSNQEELKFTREN-----------LKKIEERLKNVFIEFYRKLRHLKNYSFLNTLAISKI
AtPHO1_H8    PLSTIRNVLKLSNKEDIKFTKEN-----------LKKIEERLKNVFIEFYRKLRHLKNYSFLNTLAISKI
AtPHO1_H9    PLLTLKMMIL-GLPSEQTFSKPE-----------LRRAEELMNRAFVEFYQKLRFLKSYCFLNQLAFAKI
AtPHO1_H10   PITTLKGVFGDSNE---PISKKG-----------LKKGEEQLRLVFSEFYQKLRRLKEYSFMNLLAFSKI
Clustal Consensus    :                          :   *  :  .: *:.: * ** *  :*  *. :*
```

───────── well conserved region ─────────

```
                430       440       450       460       470       480       490
             ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AtPHO1       MKKFDKVAGQNASSTYLKVVKRSQFISSDKVVRLMDEVESIFTKHFANNDRKKAMKFLKPHQTKDSHMVT
AtPHO1_H1    LKKFDKVTGKQILPIYLKVVESSYFNISDKVMILSDEVEEWFIKHLAGENRRKAMKYLKPHHRKESHSVT
AtPHO1_H2    LKKYDKITSRNASKSYMKMVDNSYLGSSDELMKLIQRVESTFIKHFANGHRRKGMNILRPQMKREKHRVT
AtPHO1_H3    LKKYDKITSRDATKPYMKVVDSSYLGSSDEVMRLMGRVEATFIKHFANANRAKAMNILRPKAKRERHIT
AtPHO1_H4    MKKYDKIAKRNAAKLYMEMVDKSFLSSSEEVHKLLLKVESIFIEHFSNSNRREGMSHLRPKINKERHLIT
AtPHO1_H5    LKKYDKITSRHASKSYMKMIDNSYLGSSDEVTRLVERVEATFIKHFSNANRSKGMNILRPKAKRERHIT
AtPHO1_H6    LTKYDKITSRDAAKSYMKMVDKSCLGSSDEVMKLMENVEATFIKQFTNGNRTKGMNILRPKPKRERHRLT
```

```
AtPHO1_H7         MKKYDKIASRSAAKPYMEMVDKSYLTSSDEINKLMLRVESTFVEHFAGLNRSKGMNLLRPKVKKEKHRIT
AtPHO1_H8         MKKYDKIALRNAAKLYMEMVDKSYLTSSDEINKLMLRVESIFVEHFAGSNRSKGMNLLRPKVTKEKHRIT
AtPHO1_H9         LKKYDKTTSRNASKPYLNTVDHSYLGSCDEVSRLMSRVEATFIKHFANGNHREGMKCLRPKTKREKHRIT
AtPHO1_H10        MKKYEKIASRNASRNYMKIVDNSLIGSSDEVNRLLERVEVTFVKHFSSGNRREGMKCLRPKVKRERHRVT
Clustal Consensus :.*::* : :        *:: :. * :   .::: *  .**  * ::::. .: :.*. *:*:  :: * :*

                           500       510       520       530       540       550       560
                  ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AtPHO1            FFVGLFTGCFISLFVIYIILAHLSGIFTSSDQ-VSYLETVYPVFSVFALLSLHMFMYGCNLYMWKNTRIN
AtPHO1_H1         FFIGLFTGCFVALLAGYIIVAHLTGMYRQHSANTFYMETAYPVLSMFGLLFLHLFLYGCNIFMWRKARIN
AtPHO1_H2         FSTGFSAGCIFSLIVALVAIIRTRKTMPEAEH-NTYMNTMFPLYSLFGFIVLHITMYAIDIYYWKRYRVN
AtPHO1_H3         FSTGFSAGCVFSLIVALVAIIRTRNLLEMEGQ-KEYMNTMFPLYSLFGFIVLHIIVYAANIYYWRRYRVN
AtPHO1_H4         FSTGFFFGCGISLIVALGLIIHARNIMGTPGQ-RTYMETMFPLYRFFGFVVLHMDVYAANIYFWRRYRVN
AtPHO1_H5         FSTGFLGGCLFSLVVALFAIIRTRNILQEEGQ-KQYMNTMFPLYSLFGFVVLHIPMYAGNIYYWRRYRVN
AtPHO1_H6         FSTGFLGGCMFSLIVALVAIVRTRNILQDDGQ-KQYMNTMFPLYSLFGFIMLHMTMYAANIYFWRQYRVN
AtPHO1_H7         FSTGFFVGCTVSLVVALVMFIHARNIMGAVGH-KVYMETMFPLYSLFAFVVLHMIMYASNIYFWKRYRVN
AtPHO1_H8         FSTGFFVGCTVSLVIALGLFIHARNIMGAVGH-KLYMETMFPLYSLFAFVVLHMIMYASNIYFWKRYRVN
AtPHO1_H9         YFLGFFSGCAVALAIAITVLVHIRGLTKSEGR-HQYMENIFPLYSLFGFVAVHLFMYAADIYFWSRYRVN
AtPHO1_H10        FFSGFFSGCSIALVIAVVFKIESRKIMEKNYG-TEYMANIIPLYSLFGFIILHMLMYSANIYFWKRYRVN
Clustal Consensus :  *:  ** .:*        .           *: .  *:  .*.:: :*: :*. ::: * . *:*

                           570       580       590       600       610       620       630
                  ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AtPHO1            YTFIFEFAPNTALRYRDAFLMGTTFMTSVVAAMVIHLILRASGFS---ASQVDTIPGILLLIFICVLICP
AtPHO1_H1         YSFIFELGSKNELKYRDVFLICTASMSAIAGVMFVHLSLLEKGYS---FRQVQVIPGLLLLGFLLILICP
AtPHO1_H2         YAFIFGCKQGTELGYRQVLFLGFTIGTFALLCVLGNLDMEVNPKTKNFKPLTELLPLFLLVALFVALIMP
AtPHO1_H3         YSFIFGFKQGTELGYRQVLLVGFSIGVLALLCVLANLDMEADPKTKAYQARTEILPLILLAAMFIVLVLP
AtPHO1_H4         YSFIFGFKQGTELGYRHVLLLSFGLGTLSLCAVLLNLDMEMDAQTKDYRLVTELIPLFLLVLVIIIVLCP
AtPHO1_H5         YSFIFGFKHGTELGYRQVLFVGLSIGVFALLCILANLDMEVDPETKDYQALTELLPLFLLTGMFVVLVLP
AtPHO1_H6         YSFIFGFKQGTELGYKQVLFVGFSIGALALLCVLANLDMETDPKTKDYQALTELLPLFLLIAMFVVLVVP
AtPHO1_H7         YPFIFGFKEGTELGYRHVLLLSFGLGTLALCAVLINLDMEMDPNTNDYKTMTELLPMFILALVVAILFCP
AtPHO1_H8         YPFIFGFKEGTELGYGHVLLLSFGLGTLALCAVLVNMDMEMDPNTNDYKTITELVPLFVVALVIAISVCP
AtPHO1_H9         YPFIFGFEQGNDLGYREVLLVGPGLAVLTFGGVISNLDMEMDPRTKSFSVITELVPLALLVCLMMVLFCP
AtPHO1_H10        YTFIFGFKQGTELGDREVFLVSTGLAVLAFVCFLNLQLDMDWRMKHHKTLPEVIPLCLATIVLFILFCP
Clustal Consensus *.***     . *   ..:::          .. :: :  .       : :*  :  ..   . *

                           640       650       660       670       680       690       700
                  ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
```

227

```
AtPHO1           FNTFYRPTRFCFIRILRKIVCSPFYKVLMVDFFMGDQLTSQIPLLRHLETTGCYFLAQ-SFKTHEYNTCK
AtPHO1_H1        LNIFYKSSRYRLISVIRNIVFSPLYKVVMLDFFMADQLCSQVPMLRNLEYIACYYITG-SYATQDYEYCM
AtPHO1_H2        FHFLYRSTRFFFLTCLLHCLAAPLYKVTLPDFFLGDQLTSQVQALRSINFYICYYGWG-DFKKRQ-NTCE
AtPHO1_H3        FNYFYRSSRFFFLTCLFHCLAAPLYKVTLPDFFLGDQLTSQVQAIRSIEFYICYYGWG-DFRHRK-STCK
AtPHO1_H4        FNILYRSSRFFFLSVLFRCIAAPFYAVHLPDFFLGDQLTSQVQALRSLEFYICYYGFG-DFRYRRRNTCT
AtPHO1_H5        FNIFYRSSRFFFLTCLFHCLAAPLYKVTLPDFLVGDQLTSQVQALRSIQFYICHYGWG-DYKHRI-NTCT
AtPHO1_H6        FNIFYRSSRFFFLTTLFHMLAAPLYKVTLPDFFLADQLCSQAQTLRSIEFYICYYGWG-DFKQRK-NTCK
AtPHO1_H7        FNIFYRSSRVFFLMVVFRCIAAPLYKVNLPDFFLADQLTSQVQALRSLEFYICYYGWG-DFKHRQ-NTCR
AtPHO1_H8        FNIFYRSSRFFFLMVLFRCIAAPLYKVNLPDFFLADQLTSQVQALRSLEFYICYYGWG-DFKQRQ-STCK
AtPHO1_H9        FNIIYRSSRYFFVGSVFRCLLSPLYKVILPDFFLADQLTSQVQTFRSLLFYVCYYGWGGDFKRRT-HTCY
AtPHO1_H10       FNIIYRSSRFFFIRSLFHCICAPLYEVTLPDFFLGDHLTSQIQAIRSFELFICYYGLG-EYLQRQ-NKCH
Clustal Consensus  :: :*::* :: : . : :*:* * : **::.*:* **   :* :    *::   .: :   *

                            710       720       730       740       750       760       770
                   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AtPHO1           NGRYYREFAYLISFLPYFWRAMQCVRRWWDESNPDHLINMGKYVSAMVAAGVRITY-----ARE-NNDLW
AtPHO1_H1        RVKYYRDLAYAVSFLPYYWRAMQCARRWFDEGETSHLVNLGKYVSAMLAAGTKVAY-----EKE-RSLGW
AtPHO1_H2        ASEIYIYSLYIVASLPYLSRLLQCMRRMIEERSLDQGYNGVKYLLTVIAVSLRTAYGYEVKNTKNPTSHL
AtPHO1_H3        ESDVYNTFFFIVAVIPYVSRLLQCLRRLFEEKNPEQGYNGLKYFLTIVAVCLRTAY--SIQK---GQVAW
AtPHO1_H4        SNIGFRTFYFIVAVIPYWLRFLQCIRRMVEDRDLSHGYNGIKYLLTIVAASLRTAY------TLNRGSNW
AtPHO1_H5        ESDAYNAFLFIVAVIPYVSRLLQCLRRLFEEKNPEQGYNGLKYFLTIVAVCLRTTY--SVDED--NQFIW
AtPHO1_H6        DSQVFNTFLFIVSAFPFFSRFLQCMRRMLEEKNIEQGYNGFKYIVIVVAVCLGMAY--EVDDEKDRQIIW
AtPHO1_H7        SSDVYSTFYFIVAVIPYWSRFLQCVRRLIEENDSSQGYNALKYLLTVVAVCLRTAY------SFNRGNIW
AtPHO1_H8        SSDVYSTFYFIVAVIPYWSRFLQCVRRLIEEKDVSQGFNALKYLLTIVAVCLRTAF------SINRGNDW
AtPHO1_H9        DSEIYKELYLVVAIIPYWFRFAQSIRRLVEEKDKMHGLNALKYLSTILAVAARTIF------EMKRGTYW
AtPHO1_H10       SHGVYNAFYFVVAVIPYWLRFLQCIRRLCEEKESVHGYNALKYMLTIIAVIVRTAY------ELKKGRTW
Clustal Consensus       :      :: :*: * *. ** :: . : * **. ::*.     :

                            780       790       800       810       820       830       840
                   ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
AtPHO1           LTMVLVSSVVATIYQLYWDFVKDWGLLNPKSKNPWLRDNLVLRNKNFYYLSIALNLVLRVAWIETIMRFR
AtPHO1_H1        LCLVVAMSSVATIYQLYWDFVKDWGLLQHNSNNPWLRNQLMLRQKSIYYFSMVLNLVLRLAWLQTVLHSS
AtPHO1_H2        KVLAGSSSILAAVFCTYWDFVHDWGLLNKTSKNRWLRDKLLIPQKKVYFIAMILNVVLRFAWLQTILNFE
AtPHO1_H3        RVLAAVFSFIAAIFCTYWDFVHDWGLLNRTSKNRWLRDKLLVPQKKVYFIAMVLNVLLRFAWIQTVLDFN
AtPHO1_H4        NITAWVFSGVATFYGTYWDIVLDWGLLQRGCKNSFLRDKLLVPHKTVYYAAMVLNVLLRLVWLQTVLDLK
AtPHO1_H5        RILAGIFSAIAAIFCTYWDLVYDWGLLNRTSKNPWLRDKLLVPQKKVYFIAMILNILLRFAWLQTVLDFN
AtPHO1_H6        RLLGGITSAMAVVFCTYWDLVYDWGLLNRTSKNPWLRDNLLIPHKEVYVLAMILNVVLRFAWMQTVLDFK
AtPHO1_H7        KISAWVFSALATFYGTYWDIVFDWGLLHRPSK-HLLREKLLVPHKAVYYVAIVLNIVLRMAWLQTVLDFN
```

```
AtPHO1_H8        KIAAWVFSGLATFYGTYWDIVYDWGLLHRPSK-SWLREKLLVPHKSVYYVAMVVNVVLRLAWLQTVLDFN
AtPHO1_H9        LTVAVTTSSIATLFNTYWDIFRDWGLMNRNSKNPWLRDKLLVPYKSIYFIVMVANVVLRLAWMQTVLGIK
AtPHO1_H10       MILALVSSGVATGMNTFWDIVIDWGLLRKHSKNPYLRDKLLVPHKSVYFAAMVVNVILRVAWMQLVLEFN
Clustal Consensus    * :*.      :**:. ****:.  .:   **::*::  * .*    :  *::**..*:: ::
```

```
                             850       860       870       880       890       900
                     ....|....|....|....|....|....|....|....|....|....|....|....|....|...
AtPHO1           -VSPVQSHLLDFFLASLEVIRRGHWNFYRVENEHLNNVGQFRAVKTVPLPFLDRDSDG-----
AtPHO1_H1        -FEHVDYRVTGLFLAALEVIRRGQWNFYRLENEHLNNAGKFRAVKTVPLPFREVDEED-----
AtPHO1_H2        -FEFLHKQTTLAVVASLEIMRRGMWNFFRVENEHLNNVGKFRAFKSVPLPFN-YDEDDEKDD-
AtPHO1_H3        -FSFMHRQTMVAIVASLEIIRRGIWNFFRLENEHLNNVGKYRAFKSVPLPFN-YDEDDDKDN-
AtPHO1_H4        -FSFLHRETMVALMACLEIIRRGIWNFFRLENEHLNNVGRYRAFKTVPLPFN-YEEDGDHHNN
AtPHO1_H5        -FSFMHRQTMVAVVASLEIIRRGIWNFIRLENEHLNNVGKYRAFKTVPLPFN-YDEDDDKGN-
AtPHO1_H6        -FESIHTQTVVAVVASLEIIRRGIWNFFRLENEHLNNVGKYRAFKAVSLPFN-YEVDH-----
AtPHO1_H7        -LSFLHRETMIALLAALEIIRRGIWNFFRLENEHLNNVGKFRAFKSVPLPFN-YNEEEDRDS-
AtPHO1_H8        -ISFLHRETMVALIAILEIIRRGIWNFFRLENEHLNNVGKFRAFKSVPLPFN-YDEEEDRDS-
AtPHO1_H9        EAPFLHKRALVAVVASLEIVRRGIWNFFRLENEHLNNVGKYRAFKSVPLPFQELGGSKSV---
AtPHO1_H10       -LKSLHKIAVTSIISCLEIIRRGIWSFFRLENEHLNNVGKYRAFKSVPHPFHYYDDDDVDKDD
Clustal Consensus    :.        .:: **::*** *.* *:*******.*::**.*:*. **       .
```