Durham E-Theses

Characterization of the function of a novel protein tyrosine phophatase predicted by the genome sequence of Arabidopsis thaliana

Reignoux, Sabrina Natacha Aurore

CHARACTERIZATION OF THE FUNCTION OF A NOVEL PROTEIN
TYROSINE PHOPHATASE PREDICTED BY THE GENOME SEQUENCE OF
ARABIDOPSIS THALIANA.

by

Sabrina Natacha Aurore REIGNOUX

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

Master of Science

University of Durham

2003

CHARACTERIZATION OF THE FUNCTION OF A NOVEL
PROTEIN TYROSINE PHOPHATASE PREDICTED BY THE
GENOME SEQUENCE OF *ARABIDOPSIS THALIANA*.

by

Sabrina Natacha Aurore REIGNOUX
A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

University of Durham

2003

Approved by _____
Chairperson of Supervisory Committee

_____

_____

_____

Program Authorized
to Offer Degree _____

Date _____

University of Durham
ABSTRACT

IDENTIFICATION AND CHARACTERIZATION OF A NOVEL
PROTEIN TYROSINE PHOPHATASE IN HIGHER PLANTS

by Sabrina Natacha Aurore REIGNOUX
Chairperson of the Supervisory Committee: Professor J.A. Gatehouse

Department of Biological Sciences

Covalent modification of proteins by reversible phosphorylation, catalysed by specific protein kinases and phosphatases, is an important mechanism for enzyme activity regulation in metabolism and signalling pathways. Whereas both groups of proteins are well-characterized in animals, and in microorganisms, our understanding of plant protein phosphatases is limited especially concerning protein tyrosine phosphatases. For this project, genes in the model plant *Arabidopsis thaliana* potentially coding for protein tyrosine phosphatases were extracted from the genomic sequence database. From it, a dual specific phosphatase was selected. The predicted protein, called *At*PTPKIS1, appeared to contain a motif corresponding to the kinase interaction sequence (KIS) of the SNF1 family. Homologues of this gene are present in other plant species, such as tomato. The ESTs available for *At*PTPKIS1 suggested that this gene could undergo alternative splicing to form mRNAs encoding a full-length and a truncated arrangement of the protein. The predicted interaction between *At*PTPKIS1 and AKIN11, an *Arabidopsis* SnRK1, containing the motif KIS, was demonstrated both by assays using yeast two-hybrid system, and by "GST pull-down" assays *in vitro*. However, attempts to produce functional *At*PTPKIS1 as a recombinant protein in a yeast expression system for further characterisation were inconclusive. The data show that *At*PTPKIS1 can associate with SnRK1, but do not establish whether this interaction plays a role *in planta* in regulating the activity of SnRK1, which itself functions as a global metabolic regulator. *At*PTPKIS1 potential function in cells was investigated through establishing its expression pattern. Analysis of EST microarray databases showed opposite pattern of expression between AKIN β subunit and *At*PTPKIS1. These data strongly suggest that *At*PTPKIS1 is playing a role in modulating the activity of SnRK complexes. However, *Arabidopsis* mutants containing "knockouts" for this gene and similar genes showed no phenotype, and further work will be necessary to investigate the potential role of *At*PTPKIS1 in metabolic regulation.

For Darrell, my Mum and my Dad.

Table of Contents

List of figures

Acknowledgments

Abbreviations

ABA: Abscisic Acid.

ADP: Adenosine Diphosphate

ATP: Adenosine Triphosphate.

AKIN10: gene for SnRK1 in *Arabidopsis thaliana.*

AKIN11: gene for SnRK1 in *Arabidopsis thaliana.*

ALT: ALTernatively spliced.

AMP: Adenosine Mono-Phosphate.

AMPK: AMP-activated Protein Kinases.

ASC: association with SNF1 complex.

A*t*CBL1: gene for Calcineurine B like protein in *Arabidopsis thaliana.*

A*t*MPK1: gene for MAP Kinase 1 in *Arabidopsis thaliana.*

A*t*MPK4: gene for MAP Kinase 4 in *Arabidopsis thaliana.*

A*t*PTPKIS1: gene for dual specific Protein Tyrosine Phosphatase in *Arabidopsis thaliana.*

BSA: Bovine Serum Albumin.

CaMK: Mammalian Calmodulin dependent Kinase.

CBS: Cystathionin-β-Synthase.

cDNA: complementary DNA..

CDPK: calmodulin-dependent protein kinases.

D: dextrose.

DNA: DeoxyriboNucleic Acid.

dNTP: denucleotides phosphate.

DsPTP: Dual specific Protein Tyrosine Phosphatase.

DTT: Dithio Treitol.

E. coli: Esherichia coli.

EDTA: Ethylene Diamine Tetra acetic Acid

EST: Expression Sequenced Tag.

FL: Full Length.

G: Glycerol.

GST: GlutathioneS-Transferase.

HIC: Hydrophobic Interaction Chromatography.

His3: Histidine synthesis reporter gene.

HMG-CoA: HydroxyMethylGlutaryl-Coenzyme A.

HRK: HMG-CoA Reductase Kinase.

HRP: HorseRadish Peroxidase.

IPTG: IsoPropyl-1-ThioβD-Galactopyranoside.

KAPP: Kinase Protein Phosphatase PP2A.

KIS: Kinase Interaction Sequence.

LacZ: galactosidase reporter gene.

LB: Luria-Bertani medium.

LePTPKIS1: gene for dual specific Protein Tyrosine Phosphatase in tomato.

MAP: Mitogen Activated Protein.

MAPK: MAP Kinase.

MAPKK: MAPK Kinase.

MAPKK: MAPKK Kinase.

MKP1: MAPK Phosphatase 1.

MKP3: MAPK Phosphatase 3.

MKP4: MAPK Phosphatase 4.

MKP6: MAPK Phosphatase 6.

mRNA: message RiboNucleic Acid.

NFAT: Nuclear factor of activated T cells.

NR: Nitrate Reductase.

PAGE: PolyAcrylamide Gel Electrophoresis.

PCR: Polymerase Chain Reaction.

PEP: PhosphoEnol Pyruvate.

PP1: Phosphoprotein Phosphatase 1.

PP2A: Phosphoprotein Phosphatase 2A.

PP2B: Phosphoprotein Phosphatase 2B.

PPM: Phosphoprotein Phosphatase M family.

PPP: Phosphoprotein Phosphatase P family.

PTP: Protein Tyrosine Phosphatase.

r.p.m.: rotation per minute.

RLK: Regulator-Like protein Kinase.

SDS: Sodium Dodecyle Sulphate.

SNF1: Sugar Non-Fermenting 1.

SnRK: SNF1 protein Related Kinase.

SPS: Sucrose Phosphate Synthase.

TAE: Tris Acetate EDTA.

TBST: Tris Buffer Saline Tween.

TCA: Tetra Chloride Acid.

TEMED: TEtra MethylEthylenne Diamine.

X-gal: substrate galactoside.

YP: Yeast extract Peptone.

Chapter 1

1. Introduction

1.1. General introduction.

Plants respond to environmental and physiological changes, like the presence of pathogens, or lack of nutrients, or circadian rhythm, by the induction of a complex network of reactions involving both external and internal signals (Odjakova and Hadjiivanova, 2001; Martin *et al.*, 2002; López-Bucio *et al.*, 2002; Yanovsky, 2001). The responses include both the synthesis of new protein molecules, and the regulation of the activity of pre-existing proteins. For example, the presence of a pathogen would induce the production of an arsenal of defensive proteins, as well as reactions dependent on the activity of pumps, enzymes or transporters essential to defence mechanisms.

The synthesis of new protein molecules normally corresponds to the up-regulation of genes. The promoter activity of genes subject to regulation would be induced or repressed according to the needs of the cell and of the entire organism, resulting in the production of mRNA which can be translated (if induced) into a protein (O'Malley, 1977; Jacob and Monod, 1961). The activity of the proteins themselves, which may be enzymes, cytoskeletal proteins, ion channels, receptors or transcription factors, can then be modulated by reversible changes called post-translation regulation. Post-translational regulation is often due to covalent changes that alter the conformation of the protein. This change allows protein accessibility (Fig. 1.1.): for example, the catalytic site of an enzyme (Monod *et al.*, 1963) or an open channel in a transporter (Jencks, 1989). The subject of this thesis

is post-translational regulation by phosphorylation/dephosphorylation, a mechanism found throughout all living organisms. Regulation by phosphorylation is involved in both signalling pathways and metabolic pathways. For example, signal transmission *via* a phosphorylation cascade is present in the MAPK pathway, which leads to the activation of transcription factors involved in the regulation of developmental and cell division genes, in response to environmental stress or to pathogene incursion (Stone and Walker, 1995; Liu *et al.*, 2000; Yang *et al.*, 2001). In the case of regulation of metabolism, phosphorylation activates some key enzymes such as PEP carboxylase (Chollet *et al.* 1996), nitrate reductase, and sucrose-phosphate synthase SPS (Kaiser and Huber, 2001; Winter and Huber, 2000). Signal transduction mechanisms involving phosphorylation will be illustrated of example in section 1.4.

The phosphorylation/dephosphorylation regulatory mechanism depends on the action of two types of proteins (Fig. 1.2.), protein kinases (phosphorylation) and protein phosphatases (dephosphorylation). These proteins are discussed in detail in a following section 1.3. The main sites of phosphorylation in proteins are the side chain hydroxyl groups of serine, threonine and tyrosine residues, although phopsphorylation of other side chains, e.g. histidine, also occurs (Krebs and Beavo, 1979; Huang *et al.*, 1991). Eukaryotic protein kinases and phosphatases are often divided into classes which, phosphorylate or dephosphorylate proteins at either serine/threonine residues or tyrosine residues, with a further group, termed "dual-specificity" active at both serine/threonine and tyrosine (Kreis and Walker, 2000). The subject of this thesis is a protein from the model plant *Arabidopsis thaliana*, which is predicted to encode a dual-specificity protein phosphatase. The protein itself has been characterized by a "reverse genetics" approach, in which its encoding gene has been firstly identified using bioinformatic analysis then isolated and cloned as recombinant DNA. This protein also contained a domain with similarity to kinase interaction sequences, which suggested that it might form part of a regulatory system involved in global metabolic regulation.

Fig. 1.1. Enzyme activation control by conformational changes.

Reversible alteration by regulator molecule (**R**) of the three dimensional structure which influence the availability of the active site (**A**) for the substrate.

Fig. 1.2. Protein phosphorylation reaction.

Protein kinase acting as catalyser in reaction (**1**) and protein phosphatase acting as catalyser in reaction (**2**); the reaction (**1**) uses one ATP and (**2**) release inorganic phosphorus through hydrolysis.

1.2. Control of enzyme activity.

Due to changing needs within the cell, the activities of many enzymes (and other proteins with biological functions, including transporters and receptors) are continually changing. There are a number of mechanisms, which are used to regulate the activities of enzymes and other cellular proteins. The mode and the reversibility of the activation will be different depending on the enzyme and sometimes on its function.

The most basic method of regulating the activity of an enzyme is defined by its synthesis and degradation. Enzyme synthesis is regulated by control at the level of gene transcription ("transcriptional regulation"), and may also be regulated at the levels of post-transcriptional processing of RNA, and of translation of mRNA to protein ("post-transcriptional regulation"). The amount of enzyme present in the cell is also determined by degradation, and entry of the protein into the protein catabolic pathway; the half-life time is specific to the protein (Jacob and Monod, 1961). Enzymes are also regulated by processes, which do not change the amount of protein, but alter its specific activity. This processe can be reversible or irreversible, as in the case of proteolytic activation cascades (Neurath, 1986). Enzyme activity can be controlled by the presence or absence of inhibitors, the availability of the substrate or cofactor, or by product inhibition (Pardee, 1959). But the main mechanisms involve changes in covalent structure or reversible changes in conformation. In the latter case a regulator molecule binds to the enzyme and alters its spatial structure and the accessibility of the catalytic site (Fig. 1.1.) without any change in its primary structure. Change in conformation can be due to a different assembly of the subunits of a protein complex (Levine and Ginsburgh, 1985).

Covalent changes can be reversible or irreversible depending on the mode of action. In the case of reversible changes which are used in enzyme regulation, covalent changes can be an ADP-ribosylation, adenylation, or, most commonly, phosphorylation (Fig. 1.2.) (Stadtman and Chock, 1978). Other

covalent modifications of proteins, such as methylation, acetylation, palmitoylation, tyrosinolation, sulphation are not thought to play a role in regulation. In the case of protein phosphorylation, the subject of this thesis, a phosphorylate group is added to the hydroxyl groups of the side chains of serine, threonine or tyrosine residues, or (less usually) the heterocyclic basic side chain of a histidine residue, in a phospho-transfer reaction utilizing ATP as phosphate provider. The opposite reaction is a hydrolysis (Stadtman and Chock, 1978).

The best known of regulation of enzyme activity by reversible phophorylation is of glycogen phosphorylase, which hydrolyses $(glycogen)_n$ into α-D-glucose 1-phosphate and $(glycogen)_{n-1}$ in mammalian cells (Green and Cori, 1943; Cori and Cori, 1943). This enzyme is phosphorylated by a protein kinase (phosphorylase kinase) and dephosphorylated by a protein phosphatase (phosphorylase phosphatase) (Fisher and Krebs, 1955; Krebs and Fischer, 1956). The dephosphorylated enzyme requires AMP (and other ligands) for activity whereas the phosphorylated enzyme is active in the absence of AMP (Cori and Green, 1943); in effect, the phosphorylated form of the enzyme is an active form, and the dephosphorylated form relatively inactive. The reversible phosphorylation of the enzyme thus controls its activity, and through its activity, the breakdown of glycogen (Cohen, 1978; Cohen, 1983).

This modification in the covalent structure can induce allosteric conformation changes, which can block access to the catalytic domain (Johnson *et al.*, 1993) or can regulate interaction among protein partners to form a complex (Pawson, 1995). In the case of glycogen phosphorylase, the change in conformation caused by phosphorylation of serine 14 has been directly visualized by determination of structures of the phosphorylated and dephosphorylated proteins by X-ray crystallography (Sprang et al., 1988). The protein kinases and phosphatases, which are responsible for phosphorylating and dephosphorylating enzymes, and other proteins, are a large and diverse families of proteins in eukaryotes. They will be described in the following section.

1.3. Protein kinases and phosphatases in plants.

Although plants contain large families of genes encoding protein kinases and phosphatases, which have the same general functions in signalling pathways, enzyme regulation and metabolic regulation as protein kinases and phosphatases in other eukaryotes, the functions of individual genes have only been characterized in a comparatively small number of cases in plants. In particular, signalling pathways in plants are much less well understood than in higher animals, and many of the roles assigned to particular plant protein kinases and phosphatases are inferred by sequence similarity to animal proteins whose function has been characterized. Despite an overall lack of detailed knowledge, the functions of some plant protein kinases and phosphatases have been investigated (Kreis and Walker, 2000). The properties of some of these enzymes are summarized below.

1.3.1 Protein kinases.

Protein kinases in plants are divided into two types, using a biochemical definition; the protein tyrosine kinases, which phosphorylate tyrosine residues on their target proteins, and protein serine/threonine kinases, which phosphorylate serine and/or threonine residues. These proteins can also be assigned to families in a classification based on their sequence similarity. Although the biochemical definition and organization was the first aspect used to classify protein kinases, it has quickly been replaced by molecular biology and sequence analysis, and many proteins have been assigned to a particular type on the basis of sequence similarity alone (Hardie, 1999). Nevertheless, in those cases where the biochemical function has been investigated, the classification based on sequence similarity has usually proved to be valid. Protein kinases have associated protein phosphatases, and the two members of the pair show coordinated (and complementary) regulation. This is particularly the case for protein serine/threonine kinases.

1.3.1.1. Protein serine/threonine kinases

Protein serine/threonine kinases are a vast superfamily of enzymes, with the common feature that they phosphorylate serine and threonine amino acids of

some proteins, thus regulating their activities. 175 were counted in the genome of *A. thaliana*, in August 1998, 549 had been identified in higher plants in the EMBL Database (Hardie, 1999). They are grouped into 19 subfamilies, of which MAPK (mitogen-activated protein kinases), AMPK (AMP-activated protein kinases), CDPK (calmodulin-dependent protein kinases) and RLK (receptor-like kinases) are the most well known.

CDPKs have been identified in a number of plants; analysis of the *A. thaliana* genome shows that 13 members of this family are present. Different CDPKs are characterized by substrate specificity, in terms of the peptide sequence around the target residue that is recognized for phosphorylation, location (cell type) and at which developmental stage they are expressed (Lee *et al.*, 1998; Hong *et al.*, 1996). The CDPKs contain a kinase domain, similar to the one present in CaMKs (mammalian Calmodulin-dependent Kinase). This domain, corresponding to the catalytic domain, is followed by the junction region and the calcium-binding domain (Harper, 1991). RLKs in *A. thaliana* make up a group of 18 different genes. RLKs are type I membrane proteins with one main domain, which is the kinase domain. The encoded proteins contain a signal peptide at their N-terminal region, and are predicted to have an N-terminal region containing carbohydrate side chains, outside the cell. By analogy to proteins in other organisms, the N-terminal region may have receptor properties, although this has not yet been established in plants (Hardie, 1999), nor have potential ligands been characterized.

The MAP protein kinases (MAPK) are components of major signalling cascades, including most of those involved in stress responses. In plants such as tobacco and barley, auxin, abscisic acid, cold shock and wounding leaves induces the activation of MAPK (Mizozuchi T *et al.*, 1994; Knetsch *et al.*, 1996; Brogre *et al.*, 1997). These proteins also play a major part in pathogen responses where salicylic acid activates MAPK in tobacco cells (Zhang *et al*, 1997). 23 MAPKs have been identified in the *A. thaliana* genome (figures reviewed by Jonack *et al.*, 2002). The MAPK family is subdivided in 5 subfamilies based on their common signature

motif, T(E/D)Y; and designated A to D respectively plus MHK. MAPK are phosphorylated on both threonine and tyrosine residues during their activation. MAPK upstream components of the signalling pathway are refered to as MAPKK (MAPK kinases). MAPKK are dual specificity kinases capable of phosphorylating both serine/threonine and tyrosine (Garrington *et al.*, 1999; Interpro database entry IPR003527, http://www.ebi.ac.uk/interpro). The phosphorylated MAPK are inactivated by PTPs and serine/threonine protein phosphatases. In *A. thaliana*, one tyrosine specific and two dual specific protein phosphatases have been shown to inactivate *At*MPK1 and *At*MPK4 (Huang *et al.*, 2000; Ulm *et al.*, 2001; Gupta *et al.*, 1998).

1.3.1.2.   SnRKs (SNF1-related Kinases)

SnRKs are plant homologues of the AMP-activated kinases (AMPK) in mammals and SNF1 in the yeast *Saccharomyces cereviseae*. Members of this family are centrally important kinases in the cell, whose main role is metabolic regulation (Hardie, 1999). AMPKs respond to cell signalling pathways through themselves being regulated by phosphorylation/dephosphorylation.

Until DNA sequencing was used to classify genes and corresponding proteins, AMPKs were identified under a different name, which was based on their ability to phosphorylate hydroxymethylglutaryl-Coenzyme A reductase (HMG-CoA reductase). The first HMG-CoA reductase kinase was extracted and shown to phosphorylate rat liver HMG-CoA reductase in 1979 (Keith *et al.*, 1979). Later, a kinase purified from rat liver was shown to phosphorylate and inactivate both acetyl-CoA carboxylase and HMG-CoA reductase (Carling, 1987) under the positive influence of AMP, which gave the name AMP-activated kinases (AMPK) to this group of enzymes. Recent reports add to the functions of the mammalian protein; it has been shown to be involved in the inhibition of nitric oxide synthase expression (Yamamoto *et al.*, 2003), and thus plays a role in regulating the signalling processes induced by nitric oxide. Phosphorylation of HMG-CoA reductase is still used as an assay to characterize SnRKs in plants (Sugden,1999),

although another possibility to characterize them is to use the synthetic peptide SAMS (Davies *et al.*, 1989), which has been used by Mackintosh and co-workers (1992) to characterize a protein kinase in cauliflower (Mackintosh *et al.*, 1992). The AMPK homologue in yeast *Saccharomices cereviseae* is termed SNF1, for sugar non-fermenting 1. SNF1 was identified in a mutant yeast strain, which only grew on glucose base media (Gancedo, 1992). By comparison of the sequence of a subunit of the yeast SNF1, Snf1p, and the catalytic subunit of mammalian AMPK, Celenza and Carlson (1986) concluded that the proteins were derived from the same ancestral sequence, and were related to each other.

AMPK, SNF1 and SnRKs are all protein complexes. The canonical AMPK/SNF1 structure is composed of three subunits. The catalytic or α-subunit contains two domains: a kinase domain and a regulatory domain. The β-subunit (or Sip/Gal for SNF1 complex), performs the role of a scaffold for protein assembly, and contains two protein interaction domains; the kinase interaction sequence (KIS) domain, which binds to the γ-subunit, and the C-terminal ASC (association with SNF1 complex) domain, which binds to the γ-subunit. The γ-subunit itself has a regulatory role on the activity of the complex, and is formed of several CBS (Cystathionin-β-Synthase) domains. The γ-subunit has been identified in maize and *A. thaliana* (Lumbreras *et al.*, 2001), and the existence of canonical SNF1/AMPK complexes in plants has been established.

The activity of the AMPK, SNF1 and SnRK kinase complexes can be regulated by the assembly of subunits as well as by phosphorylation/dephosphorylation. Their regulation will thus depend on the availability of each component of the complex and how much their meeting and interaction is possible (Jiang *et al.*, 1996; Woods *et al.*, 1996; Lumbreras *et al.*, 2001). The interaction between metabolic regulation and activity of the complex has been clarified in the yeast *Sacchromyces cereviseae*. In this organism, a low glucose environment promotes the formation of SNF1 complexes, and thus increases the kinase activity of the protein complex (Jiang *et al.*, 1997). The mammalian AMPKs,

on the other hand, have a different regulatory mechanism, where AMP level is the signal (Ferrer *et al.*, 1987). The signalling mechanism in plants is not yet clear; the absence of cyclic AMP as a regulator in plants would seem to suggest that AMP-based mechanisms of regulating SnRKs would not be present. Nevertheless, a protein from plants, with the same biochemical properties as mammalian AMPK, and which is activated by phosphorylation through AMPK kinase and inactivated by dephosphorylation by the protein phosphatase PP2A, has been identified (Shi *et al.*, 1999). Further, the AMP/ATP ratio has been shown to affect the activity of the SnRK AKIN10 in plants (Sugden *et al.*, 1999). Other molecules clearly act as regulators of plant SnRKs; for example the metabolite glucose-6-phosphate (Toroser *et al.*, 2000) acts as an inactivator, as does the protein PRL1 (Bhalerao *et al.*, 1999).

The SNF1-related kinases are involved in regulation of metabolism through their effects on key regulatory enzymes such as HMG-CoA reductase and acetyl-CoA carboxylase. Phosphorylation of yeast acetyl-CoA carboxylase by SNF1 has been detected under normal physiological conditions (Woods *et al.*, 1994), whereas, regulation of HMG-CoA reductase by the cauliflower SnRK (HRK) is only observed *in vitro* (Dale *et al.*, 1995). However, the metabolic regulatory activity of plant SnRKs has been confirmed (Sugden *et al.*, 1999) by work in which key metabolic enzymes such as nitrate reductase and sucrose phosphate synthase have been shown to be regulated by spinach SnRK. Moreover, a link has been established between potato SnRKs and the expression of sucrose synthase (Purcell *et al.*, 1998). SNF1 also regulates a variety of processes in yeast, such as sporulation, heat shock response and the cell cycle (Thompson *et al.*, 1991), but it is not known whether this is also the case for SnRKs in plants.

1.3.2.   Protein phosphatases.

Protein phosphatases catalyse the removal of phosphate groups from specific phosphorylated amino acid side chains, thus changing their covalent structure and altering the functional properties of proteins containing these residues (Johnson et

al., 1993). Protein phosphatases were originally divided into families based on their specificity of hydrolysis of phosphorylated amino acid residues; this classification gave three families, protein serine/threonine phosphatases (specific for phosphoserine and phosphothreonine residues), protein tyrosine phosphatases (specific for phosphotyrosine residues) and dual-specificity protein phosphatases (able to hydrolyse both phosphoserine/threonine and phosphotyrosine residues). This old classification is still in use, but has been largely superseded by a new classification based on sequence and structural analysis (Smith and Walker, 1996; Luan, 1998). Although the ability of protein phosphatases to dephosphorylate residues in proteins is determined by their primary specificity, a given serine/threonine phosphatase will not hydrolyze any phosphoserine residue in any protein; rather there are secondary specificity determinants based on the environment of the phosphorylated residue in the protein, and the ability of the phosphatase to interact with the potential protein substrate.

1.3.2.1.   Protein ser/thr phosphatases.

The new classification based on the sequence and structure similarity separates protein phosphatases with specificity towards phosphoserine/threonine residues into two distinct families: PPP representing PP1, PP2A, PP2B, and the PPM family where only one subfamily, PP2C, is represented (Barford, 1996; Cohen, 1997). While all are strictly specific for phosphoserine and threonine residues, two types were differentiated on the basis of their substrate specificity and their pharmaceutical properties. Type 1 protein serine/threonine phosphatases, corresponding to PP1 dephosphorylate the α subunit of phosphorylase kinase whereas type 2 phosphatases (PP2A, PP2A, PP2C) dephosphorylate the β subunit of phosphorylase kinase. Only PP1 and PP2A are inactivated by the okadaic acid and calyculin A (Li and Casida, 1992).

Studies of the three dimensional structure of proteins belonging to the PPP family revealed a common configuration in their catalytic domain, consisting of a combination of helix/sheets, which contain metal ions in the case of PP1 and

PP2B (Elgloff *et al.*, 1995). Except for PP1, which contains two subunits, the other members of the family are heterotrimers (Faux and Scott, 1996; Wera and Hemmings, 1995). PP2B is active when its three subunits (regulatory, catalytic and Ca^{2+} binding) are combined under conditions where there is a high concentration level of Ca^{2+}, and thus it is a calmodulin-dependent protein phosphatase (Klee *et al.*, 1988). In animals PP2C consists of a catalytic domain of two anti-parallel sheets in the N-terminal region attached to an anti-parallel helix in the C-terminal region (Das *et al.*, 1996). The catalytic domain contains two Mn^{2+} ions.

Protein serine/threonine phosphatases have been implicated to have functional roles in a wide variety of cellular processes. In animals, PP1 is required to regulate enzymes of glycogen metabolism such as glycogen phosphorylase, phosphorylase kinase, and glycogen synthase. Thus PP1 is involved in muscle contraction and neuronal activities (Cohen, 1989). PP1 also functions in control of the cell cycle, particularly in early embryonic development (Fernandez *et al.*, 1992), although other phosphatases, such as PP2A are also involved. In plants, both PP1 and PP2A enzymes are involved in regulation of sucrose phosphate synthase and nitrate reductase, which are key regulatory enzymes of carbon and of nitrogen metabolism. They also regulate K^{+} ion channels (Li *et al.*, 1994). Further roles for these enzymes in plants are suggested by observations that the induction of gene expression caused by ethylene and pathogen elicitors can be mimicked by okadaic acid (Raz and Fluhr, 1993; Mackintosh *et al.*, 1994), an inhibitor of PP1 and PP2A. These phosphatases are therefore involved as negative regulators of both signalling processes, since inhibition allows the response to occur in the absence of the normal stimulus. One of the targets for dephosphorylation by these enzymes *in vivo* may be a MAPK involved in cell division, which is activated by pathogen elicitors (Suzuki and Shinshi, 1995).

PP2B, which is a calmodulin-dependent protein phosphatase first extracted from brain tissues, has been shown to have an important impact on the cell signalling involved in T cell proliferation by regulating the transcriptional

factor NFAT (Loh *et al.*, 1996). PP2B is involved in regulation of K$^+$ ion channels, and possibly other ion channels also (Luan *et al.*, 1993; Allen and Sanders 1995; Bethke and Jones, 1997). In yeast PP2B plays a role in salt tolerance, which has been proved as well in *A. thaliana* (Liu and Zhu, 1998). More evidence of PP2B's function in stress signal transduction has been given with the study of *At*CBL1 expression in response to cold, drought and wounding (Kudla *et al.*, 1999).

PP2C in mammals and yeast acts in stress response by regulation of MAPK-based signalling pathways. More recently PP2C have been proved to have the same role in plants (Luan, 2002); for example, the plant PP2C-type phosphatases ABI1 and ABI2 are involved in ABA signal transduction by regulation of the MAPK pathway involved in stomatal opening and closure. Another plant PP2C called KAPP is thought to regulate a protein kinase termed RLK (regulator-like protein kinase; Meskiene *et al.* 1998), which plays a role in a number of important signal transduction pathways, for example the development of shoot meristem.

1.3.2.2.   Protein   tyrosine   phosphatases   and   dual-specificity phosphatases.

The protein tyrosine phosphatases (PTPs) are a family of enzymes able to dephosphorylate phosphotyrosine residues in proteins, or in the case of dual-specificity phosphatases, all residues (tyrosine, serine and threonine) phosphorylated on –OH groups (Denu *et al.*, 1996). The classification, based on sequence similarity in the catalytic domain, separates the family into two groups: the tyrosine specific PTPs and the dual specificity PTPs. A signature motif highly similar for all members of the family is located in the catalytic region and corresponds to the following sequence: (V/I)HCxxAGxxGR(S/T)G, where 'x' is any amino acid. The substrate specificity is not determined by the catalytic domain (Denu *et al.*, 1996).

The catalytic mechanism involves two amino acid residues in the catalytic site, cysteine and arginine, shown in bold H**C**xxAGxx**G**R. This is a two step

mechanism. The first step involves attack by the cysteine –SH group (ionised to – S- due to the neighbouring arginine residue) on the phosphorus atom in the phosphoryl group of the substrate. The phosphoryl group is transferred to the enzyme (attached to the cysteine –S) and the dephosphorylated substrate is released. The second step involves hydrolysis of the phospho-enzyme intermediate by a water molecule. The arginine residue not only helps the cysteine residue to ionise, but also plays the role of stabilization of the transition states. In addition an aspartic acid residue plays the role of acid in the first step, by donating a proton in the first step (allowing the dephosphorylated substrate to be liberated) and of base in the second step, accepting a proton from a water molecule (Fauman and Saper, 1996) to produce the hydroxyl ion that hydrolyses the phosphoenzyme intermediate. Study of the three dimensional structures of PTPs by X-ray crystallography has helped in understanding their catalytic mechanism in more detail (Denu and Dixon, 1998).

Protein specific phosphatases are classified into two subgroups based on their cellular location: transmembrane and non-transmembrane, also referred to as "receptor-like" or "non-receptor" (Fauman and Saper, 1996). The specificity of the protein phosphatase towards potential substrates is determined by regions distinct from the catalytic site. Many PTPs contain other domains, which have been shown to be involved in protein-protein interactions, and can give an idea of the possible function of the protein. For example, binding domains such as immunoglobulin or fibronectin III can be found in PTPs from mammalian cells (Gebbink *et al.*, 1993; Jiang *et al.*, 1993).

Whereas the receptor-like PTPs in mammalian cells have roles in signal transduction that are becoming increasingly well understood, the roles of many non-receptor PTPs remain obscure. In mammals a non-receptor PTP called PTP1B has been well studied (Shifrin *et al.*, 1997). PTP1B appears to activate MAPK during mitosis, and in response to stresses like osmotic shock. Dual specificity protein phosphatases have been shown to have high specificity to

MAPK (Camps *et al.*, 2000). Knowledge about the function of PTPs in general in higher plants is still poor. The existence of PTPs in *A. thaliana* and in general in higher plants has only been shown in the last few years (Xu *et al.*, 1998; Fordham-Skelton *et al.*, 1999) and the extensive families of these proteins found in mammals are not present. Nevertheless two plant dual specific tyrosine phosphatases regulate MAPK, with one involved in genotoxic stress responses (Gupta *et al.*, 1998; Ulm *et al.*, 2002)

1.4. Phosphorylation in the cell: regulation of both development and metabolism.

Metabolism and development are both subject to regulation, which is governed in both cases by cell signalling. Metabolism can be directly regulated by posttranscriptional activation of enzymes, or by the synthesis of new enzymes and other proteins. This *de novo* protein synthesis results from gene activation at the transcriptional level, induced by transcription factors activated in response to signal transduction. Cell division and development are regulated through signal transduction in response to stress, modifications in metabolism, and to the biological clock, such as circadian rhythm, which corresponds to a 20 to 28 hours cycle (Li and Xu, 2000; Martin *et al.*, 2002; Yvanovsky, 2001). In principle a first signal, which is sensed by the cell *via* a receptor, will induce a series of reactions called signal transduction (Fig 1.3.). Consequently the process of signal amplification will lead to the transcription of one or a group of particular genes into the corresponding proteins, which will have biological functions in order to answer the needs of the cell. An example of such process is the biological arsenal deployed against pathogens (e.g. inhibitors, lysis enzymes) by cells, which respond to an extracellular elicitor molecule (Kroj *et al.*, 2003). Plants encounters four types of signal transduction, which form a complex network involving combinations of molecules like enzymes, metabolites and ions. These different signal transduction categories can be defined by a series of reactions involving protein activation cascades. The series of activation reactions can involve protein kinases such as

Fig. 1.3. Phosphorylation in the cell.

Red arrows represent reversible protein phosphorylation.

Cellular signalling is represented on the upper part of the figure where a molecule (hormone, ions, metabolites) is recognized by a protein receptor (sometimes acting as a transporter) which can be either auto-activated or can activate a second messenger. The second messenger activates a first protein kinase and induces a series of phosphorylation reactions, which amplify the signal. The last phosphorylation activates the relevant transcription factor (not represented in this diagram) for the transcription of a gene into mRNA and the synthesis of the corresponding protein.

The activity of the protein synthesised can be regulated by reversible phosphorylation, which is catalysed by protein kinase or phosphatase.

Finally, the common example of protein kinase activation by ligand binding, is represented on the bottom right of the figure.

MAPK or AMPK activation, but also the sugar derivative phophoinositol, or the level of Ca^{2+} in the cell.

The phosphoinositol signal pathway is complementary to Ca^{2+} signal transduction, involving proteins such as the signal receptor, G-protein complex, Ca^{2+} ion channels, and calmodulin-dependent protein kinases and phosphatases. Calmodulin-dependent protein kinases and phosphatases are notably involved in ABA signalling. PP1, PP2A, PP2B and PP2C are protein phosphatases involved in drought-induced stomatal closure in guard cells, a process, which uses ABA as a signalling molecule (Luan, 1998).

MAPK cascade reactions are known as components of signal transduction pathways in the regulation of development and cell division, and in responses to pathogens, heat shock, wounding and high salt environment, as well as genotoxic stress. The dual specific phosphatase MKP1 is involved in salt and genotoxic stress signal transduction through negative regulation of MPK3, MKP4 and MPK6 (Ulm *et al.*, 2002). Signalling pathways involving MPK3 and MPK6 have been entirely dissected (reviewed by Jonak *et al.*, 2002). This is one of the three pathways thought to transduce pathogen signals in *A. thaliana*. A diagram is given in figure 1.4. to explain the interaction between enzymes, and how the signal, transmitted by perception of the peptide elicitor of the bacterial pathogen, is translated by the cell into a biological arsenal for resistance.

The metabolic regulator kinases AMPK, or SnRK in plants, are more concerned with the regulation of metabolism through postranscriptional activation or inactivation of key regulatory enzymes, but also through signalling pathways which result in induction of a group of enzymes which produce a metabolic response, such as enzymes of carbohydrate metabolism whose activity is regulated by sugar levels (Yu, 1999). Upstream regulation of SnRKs by phosphorylation/dephosphorylation is well known. For example, SnRK1 in plants is inactivated through hydrolysis of a phosphorylated threonine residue by PP2C,

Fig. 1.4. MAPK signalling transduction in plant response to pathogen.

The elicitor flg22 is perceived by the receptor FLS2, which activates the MAPKKK: MEKK1 and induce the phosphorylation cascade. The cascade result in the activation of the transcription factors WRKY22 and 29, which are responsible for the induction of genes of resistance to pathogens. This effect can be cancelled by inactivation of the MAPK: MPK3 and 6, by the dual specific phosphatase MKP1.

similar to a mechanism of regulation shown for mammalian AMPK (Sugden *et al.*, 1999). Not only glucose metabolism would be regulated by this kinase, but nitrogen as well. NR (nitrate reductase), which is regulated by SnRK, is one of the key enzymes for assimilation of nitrogen.

It is now accepted that nitrogen co-act as signal beside glucose, for both metabolism and development regulation. For example, glucose induces the metabolic pathway resulting in the synthesis of gibberelin, which promotes cell expansion and division, and is responsible for the difference between tall and dwarf varieties of many plant species. A direct developmental effect of nitrate as a signal has been observed for the growth of roots (Zhang *et al.*, 1998).

1.5. Aims of the project.

During the year of the associated research programme I looked at the protein tyrosine phosphatase (PTP) family in plants, in order to identify one of them and to try to elucidate its function in the cell.

The main plant model used was *A. thaliana*, for which a complete genome sequence is available. Genome data are organized and updated with new information about gene transcripts and potential function on the MIPS database (http://mips.gsf.de/proj/thal/db/search/search_frame.html). I aimed to use this database as a source to allow all putative PTPs in *Arabidopsis* to be found by a bioinformatics approach.

After sequence analysis based on alignment using PSI-BLAST (http://www.ncbi.nlm.nih.gov/blast/) and searching for sequence motifs, the programme aimed to select one putative PTP and investigate its functional properties and cellular role. Simultaneously, comparison with homologues in other species would be undertaken.

To investigate the functional properties of the selected PTP, it was proposed to use recombinant DNA techniques to produce the protein in an

expression system. Several constructs were prepared to address different questions. One was made in order to characterize the specific property of dsPTPs as phosphatases with activity towards proteins and small molecules through the production of a functional enzyme. The second plasmid construct considered was to detect the interaction with other proteins in a putative signal transduction pathway.

The last point of this research aimed to investigate the potential function of this dsPTP in the cell. A bioinformatic approach was again taken, involving analysis of publicly available expression data from DNA microarrays, using the database http://genome-www5.stanford.edu.edu/ and http://www.prl.msu.edu/circl. Analysis of expression patterns for the selected PTP was compared with those of other genes affected by circadian rhythm (cycle of 20 to 28 hours) and diurnal rhythm (day/night cycle). Further comparisons were carried out with genes implicated in biological processes such as development and genes of known location.

Chapter 2

2. Materials and methods

2.1 Materials.

2.1.1.   Reagents and chemicals.

Unless otherwise stated chemicals were provided by Sigma® (Sigma Chemical Co. Poole, Dorset, UK) or BDH (Poole, Dorset, UK).

2.1.2.   Buffers, solutions and microbial growth media.

- **1 x TBST (Tris-Buffered Saline Tween®).**

25 mM Tris-HCl, pH 7.5

0.15 M NaCl

0.05 % Tween®-20

- **TAE buffer for agarose gel electrophoresis.**

40 mM Tris-acetate, pH 8.0

0.1mM EDTA

- **6 x DNA gel loading buffer.**

0.25 % (w/v) orange G

15 % (w/v) Ficoll (type 400)

- **Glutathione 10 mM elution buffer.**

50 mM Tris-HCl, pH 8.0

10 mM reduced glutathione

- **GST pull-down lysis buffer.**

20 mM Tris-HCl, pH 8.0

200 mM NaCl

1 mM EDTA

0.5 % (v/v) NP-40

- **(his$_6$)-tag protein purification buffers.**

Loading buffer:

50 mM NaH$_2$PO$_4$, pH 8.0

300mM NaCl

5 mM imidazole

Washing buffer:

50 mM NaH$_2$PO$_4$, pH 8.0

300mM NaCl

20 mM imidazole

Eluting buffer:

50 mM NaH$_2$PO$_4$, pH 8.0

300mM NaCl

300 mM imidazole

- **2 x SDS PAGE gel loading buffer.**

As described in Sambrook and Russell (2001);

100 mM Tris-HCl, pH 6.8

4 % (w/v) SDS

0.2 % (w/v) bromophenol blue

20 % (v/v) glycerol

200 mM β-mercapto-ethanol or DTT

- **4 x SDS PAGE resolving gel casting buffer.**

1.5 M Tris-HCl, pH 9.0

0.4 % (v/v) TEMED

0.4 % (w/v) SDS

- **1 x SDS PAGE stacking gel casting buffer.**

0.12 M Tris-HCl, pH 6.8

0.10 % (v/v) TEMED

0.10 % (w/v) SDS

- **Ammonium persulphate.**

0.1 g/ml Ammonium Persulphate

- **10 x SDS PAGE runing buffer.**

0.25 M Tris

1.92 M glycine

1 % (w/v) SDS

pH 8.3

- **Coomassie stain.**

1 g/l Kenacid blue

40 % methanol

10 % acetic acid

- **Destain**

30 % methanol

10 % acetic acid

- **Western blot transfer buffer**

0.02 M Tris

0.15 M Glycine

1 mM SDS

20 % methanol

- **Western blot blocking buffer**

1x TBST, pH 7.6 (composition as above)

25 mg/ml BSA

- **Western blot sera buffer**

1x TBST, pH 7.6 (composition as above)

- **Luria-Bertani medium (low salt for zeocin antibiotic selection)**

10 g/l tryptone

5 g/l yeast extract

5 g/l NaCl

pH 7.5 with NaOH before autoclaving

For solid media 15 g/l bactoagar was added prior to autoclaving

- ### YP D/ G (Yeast Extract Peptone Dextrose/Glycerol)

10 g/l yeast extract

20 g/l peptone

20 g/l dextrose or glycerol

pH 7.5 with NaOH before autoclaving

For solid media 20 g/l bactoagar added prior to autoclave

- ### Murashige & Skoog base medium

2.2 g/l Murashige and Skoog base (Duchefa; Melford laboratories Ltd,

Chelsworth, Ipswich)

10 g/l sucrose

3 g/l phytogel

pH 5.6

2.1.2.   Plant materials.

Seeds of *A. thaliana* variety Columbia were obtained from Syngenta plc (Basle, Switzerland). Seeds were germinated on Murashige & Skoog base medium in petri dishes under plant growth room conditions (25°C, 16h light : 8 h dark). When plantlets had developed a good root system they were transferred to pots containing a 1:10 sand:compost mixture, and grown on in a glasshouse. Seeds of *Arabidopsis* knockout mutant lines were obtained from Syngenta, and were grown as described above.

2.1.3.   Molecular biology kits and reagents.

2.1.3.1.   DNA & protein size markers.

DNA size markers (λ phage DNA / *Eco* 47 I digest) were obtained from MBI Fermentas (Helena Biosciences, Sunderland, UK).

Protein size markers for SDS PAGE were the SDS7 range (Sigma®) containing a mixture of the following proteins: BSA (66 Kda); egg albumin (45 Kda);

glyceraldehyde-3-phosphate dehydrogenase (36 Kda); carbonic anhydrase bovine (29 Kda); trypsinogen (24 Kda); soybean trypsin inhibitor (20 Kda); alpha lactalbumin (14 Kda).

2.1.3.2.   Enzymes and special biological reagents.

Restriction enzymes were supplied by Promega (Southampton, UK), MBI Fermentas and Roche (Lewes, East Sussex, UK). T4 DNA ligase and "EXPAND" DNA polymerase mix were from Roche. [^{35}S]-Methionine was supplied by ICN (ICN Pharmaceuticals Ltd; Basingstoke, UK), glutathione-agarose and reduced glutathione were from   Sigma$^{®}$. Nucleotides were supplied by Promega. Oligonucleotide primers for PCR DNA amplifications were supplied by Sigma GENOSYS or MWG Biotech.

2.1.5.   DNA clones and vectors

2.1.5.1   DNA clones

The *Arabidopsis* cDNA clones encoding *At*PTPKIS1 used in this thesis were obtained   as   a   series   of   full-length   ESTs   from   ABRC (http://aims.cps.msu.edu/aims/index.html)   and   KAZUSA (http://www.kazusa.or.jp/en/plant/); The *At*PTPKIS1 transcript corresponds to partial cDNA identified as ESTs SQ104e10F, SQ163c03F and SQ091c11F. The ESTs were completely sequenced to give a composite sequence for both the "full length" (FL) and alternatively spliced (ALT) *At*PTPKIS1 mRNAs.

The cDNA encoding *Arabidopsis* SnRK AKIN11, in pET28a, was kindly provided by Dr Victoria Lumberas (Lumbreras *et al.*, 2001), CID-CSIC, Barcelona, Spain.

2.1.5.2.   Vectors

Proteins expressed as GST-fusions in *E. coli* used pGEX.5X.1 from Amersham Pharmacia Biotech (Amersham Biosciences, Little Chalfont, Buckinghamshire, UK). Expression of protein in *Pichia pastoris* utilised a modified form of the vector

pGAPZαA (Invitrogen Ltd; Paisley, UK), transformed by replacing the *Eco* RI site within the multiple cloning site with a *Nco* I site (gift of Dr Mark Skipsey, Biological Sciences, University of Durham).

2.1.6. Microbial host strains.

2.1.6.1. *E. coli* TOP 10.

One Shot® TOP 10 chemically competent cells were from Invitrogen. The full genotype of these strains can be found in the technical support book or directly on-line: www.invitrogen.com.

2.1.6.2. *Pichia pastoris* X33.

X33 is a wild type *Pichia* strain and was obtained from Invitrogen.

2.2. Methods.

2.2.1. General DNA manipulation.

2.2.1.1. Polymerase Chain Reaction.

Polymerase chain reaction (PCR) was used to amplify DNA sequences for analysis and preparation of DNA fragments for expression constructs. Two types of thermo-resistant DNA polymerases were used; analytical PCR used the non-proof reading *Taq* DNA polymerase, and PCR for expression constructs used a commercial mixture of *Taq* polymerase with a proof reading DNA polymerase ("Expand"; Roche).

PCR reactions (100 µl) contained 1 pg to 1 µg of template DNA, 1 µM of each primer, 200 µM of dNTP nucleotide mix and 1 U of DNA polymerase (added last) with the supplier's reaction buffer.

Standard PCR cycle corresponding to a denaturation of the template DNA: 30s at 94°C, followed by primers annealing: 30s at 50°C and DNA extension: 2 min, were used. (if "Expand" was used, 1 min. if *Taq* was used) at 72°C. This cycle was repeated 15 times.

PCR products were separated on 1 % agarose gels, extracted using Ultrafree®- DA spin columns (Millipore) and cloned directly into the PCR product cloning vector pCR2.1 using TA cloning (Invitrogen).

2.2.1.2. Plasmid preparations.

Plasmid DNA mini-preparation used the Wizard® *plus* SV Minipreps DNA Purification System (Promega), and all operations were carried out as described in the manufacturer's handbook. Plasmid isolation is based on the principle of alkaline SDS lysis, using silica spin column purification. A Sigma® GenElute™ Plasmid Midi Prep Kit (based on the same alkaline lysis principle) was used to

obtain larger quantities of DNA. Again, the manufacturer's procedures were followed.

2.2.1.3.   Restriction digests.

DNA was digested with restriction enzymes according to standard procedures, as described in Sambrook and Russell (2001), using the appropriate buffers as recommended by the enzyme suppliers. DNA was incubated with enzyme (at least 2 units per μg DNA) for at least 1h at 37°C.

2.2.1.4.   Agarose gel electrophoresis.

DNA was analysed by gel electrophoresis using 1 % (w/v) agarose gels, in 1 x TAE buffer containing 266 μg/ml ethidium bromide. Methods were as described in Sambrook and Russell (2001). Gels were run normally at 10 V/cm (length).

2.2.1.5.   DNA purification from agarose gel.

DNA fragments separated by agarose gel electrophoresis were visualised on a UV transilluminator (300 nm); DNA fragments were excised from the gel using a clean scalpel. DNA was then extracted from the agarose gel slice using spin columns (Ultrafree®-DA, Millipore; Amicon® Bioseparations, Bedford, USA). The gel slice was centrifuged (10 min at 5,000 x g) using a "bench top" microcentrifuge (EBA12 from Hettich).

2.2.1.6.   Ethanol precipitation of DNA.

To DNA solutions, 10 µg of glycogen was added followed by one tenth volume of 3 M sodium acetate, pH 5.2 and three volumes of ethanol. The mixture was incubated over night at -20°C and the DNA recovered by centrifuging for 10 min. at 10,000 x g using a Beckman Avanti™ 30 centrifuge and a F2402 as rotor from the same manufacturer. The DNA – glycogen pellet was washed twice with 70 % ethanol, air dried for 10 min, and resuspended in sterile water.

2.2.1.7.   Ligation of DNA.

DNA ligations contained 1 U of T4 DNA ligase in ligase buffer (as supplied by the manufacturer) in a total volume of 10 - 20 µl. Ligations were incubated overnight at room temperature. PCR products generated using "EXPAND" DNA polymerase mix were ligated into pCR2.1-TOPO (Invitrogen) according to the manufacturers instructions.

2.2.1.8.   Transformation of *E. coli*.

Transformation of plasmids into chemically competent TOP10 and BL21 cells was carried out using the "heat shock" method as described by the supplier. Transformants were selected by plating on LB agar plates containing the appropriate antibiotic.

2.2.1.9.   DNA Sequencing.

DNA sequencing was carried out on plasmid DNA purified by standard mini-preparation protocols. The sequencing methods were based on the dideoxytermination method, modified for use with fluorescent nucleotide labels. DNA sequencing reactions, and data collection, was performed by automated methods using Applied Biosystems Model 373 and 377 DNA sequencers by the DNA Sequencing Unit, Department of Biological Sciences, University of Durham. Analysis of DNA sequence data was carried out using Sequencher (Genecodes, USA) software running on Apple Macintosh computers.

2.2.2.   General protein manipulation procedures.

2.2.2.1.   Acetone precipitation of proteins.

Proteins were precipitated by addition of cold (-20°C) acetone to a final concentration of 80 % followed by over night incubation at –20 °C. Proteins were recovered by spinning for 10 min at 10,000 x g (Beckman Avanti™ 30 centrifuge with a F2402 rotor), washed in 80 % acetone. Dried pellets were resuspended in 1 x loading buffer prior to SDS PAGE.

2.2.2.2.   SDS PAGE.

Proteins were analysed under denaturing and reducing conditions by gel electrophoresis on polyacrylamide gels, using a method adapted from that described by Laemmli (1970). 12 % (w/v) polyacrylamide separating gels were prepared by mixing the following ingredients:

1.87 ml of 4 x Resolving buffer;

3.23 ml of water;

2.35 ml of 40 % polyacrylamide;

50 μL of ammonium persulphate solution (0.1 g/ml).

The mix was poured between 2 vertical spacer plates and overlaid with water before polymerisation. After the separating gel had polymerised, the water was removed, and a stacking gel layer (4% polyacrylamide) was added containing the following:

4.5 ml of 1.1 x stacking buffer;

0.5 ml of 40 % polyacrylamide;

25 μl of ammonium persulfate solution (0.1 g/ml).

Sample wells were cast in the stacking gel by insertion of a suitable comb, which was removed after the gel had set. Protein samples (30 μl) in 1 x SDS loading buffer were prepared by adding 1 μl of β-mercaptoethanol, boiled for 10 min and

spun for 30 s at 10, 000 xg prior to loading. Electrophoresis was carried out for 4 H at 100 V.

After electrophoresis, proteins were visualized by staining with Coomassie blue, followed by destaining, or were transferred to nitrocellulose for immunodetection (see below).

2.2.2.3.   Protein transfer to nitrocellulose membranes.

SDS PAGE gels were incubated in transfer buffer for 15 min prior to transfer. Nitrocellulose membranes (Type BA85; Schleicher&Schuell) were soaked in water for 5 min. A transfer sandwich was assembled in a flat plate electroblotter (ATTO Genetic Research Instrumentation Ltd; Felsted, Dunmow, UK) as described in figure 2.1. Blotting was carried out at 0.025 A for 1 H. After transfer, the blot was removed, washed and either air-dried for storage, or transferred immediately to blocking solution for immunodetection.

2.2.2.4.   Detection of histidine-tagged proteins on nitrocellulose membranes.

This technique was used for the detection of recombinant poly-histidine-tagged fusion proteins, based on the affinity of Ni^{2+} for poly-histidine tags, using INDIA™ HisProbe-HRP (a nickel-peroxidase conjugate; obtained from PIERCE; Tattenhall, Cheshire UK). Detection of histidine-tagged proteins was carried out as described by the manufacturer. The procedure consisted of an incubation of the blot with agitation at 200 r.p.m. in 10 ml of a BSA-TBST blocking buffer for 1 H, washing in TBST, followed by an incubation for 1 H at room temperature (agitated at 200 r.p.m.) in 10 ml of a 1:5000 dilution of the INDIA™ HisProb HRP reagent in TBST. The blot was again washed, and then, bound reagent was detected by covering the membrane with 1 ml ECL® mix (Amersham), and processed according to the manufacturers instructions. The blot was then exposed to X-Ray film (Fujifilm medical super RX 100NIF 18 x 24 cm) for the required time (1-5 min usually; determined empirically for each blot).

Fig. 2.1. Protein transfer onto nitrocellulose membrane.

Nitrocellulose membrane (orange) and protein electrophoresis gel (blue) are placed between three filter papers and one acetate foil (pink); all are cut to the same size as the protein gel electrophoresis; - indicates the anode where the current applied is negative; and + the cathode where the current applied is positive.

2.2.3.   Expression of Recombinant Proteins.

2.2.3.1.   GST-fusion protein expression in *E. coli*.

Coding sequences for assembly of GST-fusion protein expression constructs were amplified from plasmid DNA sources in a PCR reaction carried out under the conditions described in section 2.2.1.1. Constructs were designed to include *Eco*RI and *Xho*I restriction sites at the 5' and 3' ends, respectively, for subsequent cloning 'in frame' C-terminal to the GST coding sequence contained in the GST fusion expression vector pGEX 5X.1. A stop codon (TAG) was included in the antisense *Xho*I site containing primers. Primers and templates used were as follows:

Full length *At*PTPKIS1.

Template EST SQ104e10F.

Sense primer: $^{5'}$ GAATTCATGAATTGTCTTCAGAATCTTCC $^{3'}$

Antisense primer: $^{5'}$ CTCGAGTCAAACTTCTGCCTCAGAACAAG $^{3'}$

Alternatively spliced *At*PTPKIS1.

EST 106E24T7.

Sense primer: $^{5'}$ GAATTCATGTTGACAAGCTTCGTAAAATTGG $^{3'}$

Antisense primer: $^{5'}$ CTCGAGCTATATGTTAAGAGACAGAGAC $^{3'}$

KIS domain.

Template EST SQ104e10F.

Sense primer: $^{5'}$ GAATTCTTCAGGAAGACTGTTACTCTGACACTG $^{3'}$

Antisense primer: $^{5'}$ CTCGAGTCAGTCCACTACTTTAGCGTAATTG $^{3'}$

After amplification, the PCR the products were analysed by gel electrophoresis. The required band was excised and the DNA fragment purified as described in section 2.2.1.5. . The purified PCR product was then ligated into the intermediate cloning vector pCR2.1 using the TOPO-TA cloning method, as described by the manufacturer (Invitrogen). *E. coli* TOP10 cells were then transformed with the ligated recombinant plasmids by heat shock as described in section 2.2.1.8., and

cells were plated out on selective media (LB plates, 50 µg/ml ampicillin). Putative transformant colonies were picked off, and plasmid DNA was prepared by a standard mini-preparation. DNA was analysed by restriction digestion to confirm the presence of the desired PCR fragment as an insert. DNA from clones containing the correct insert was subjected to DNA sequencing to confirm that the correct insert was present and that no sequence errors resulting from PCR were present. This procedure allowed a specific clone to be validated, and this was used as a stock for subsequent experiments.

DNA was prepared from the validated clone by a plasmid mini-preparation (using 10 ml of bacterial culture grown over-night at 37 °C). 20 µl of DNA was digested to completion using 1 U of the appropriate restriction enzymes (*Eco*RI and *Xho*I) to produce an intact insert fragment. The fragment was separated by agarose gel electrophoresis, extracted, and purified as described in section 2.2.1.5. pGEX expression vector was restricted with the same enzymes used to extract the coding sequence from the intermediate vector, and gel purified to remove the small DNA fragment from the multi-purpose cloning site. The purified insert and expression vector were then ligated together under standard conditions (section 2.2.1.7).

3 µl of the ligation mixture was used to transform *Escherichia coli* TOP10 cells, which were plated out on selective media (containing ampicilin at 50 µg/ml). Ampicillin resistant colonies were picked off to prepare plasmid DNA by standard mini-preparation. Plasmid DNA was restricted with the appropriate enzymes, and analysed by agarose gel electrophoresis. One of the clones, that gave a band of a correct size, was then subjected to DNA sequencing to give a validated final expression clone. This was stored as 1 ml culture aliquots in LB medium containing 20 % glycerol at –80 °C. TOP10 cells could be used as expression host for this system.

2.2.3.2. Protein expression in *Pichia pastoris*

Coding sequence corresponding to the proteins required for expression in *Pichia* were obtained from expression constructs used in the yeast two hybrid system as described by Fordham-Skelton *et al.* (2002). Coding sequences in plasmid pACT2 (Clontech) were isolated as *Nco*I / *Xho*I restriction fragments and ligated into the *Nco*I / *Sal*I sites of a modified pGAPZαA *Pichia* expression vector, constructed by Dr Mark Skipsey (University of Durham). Note: the *Nco*I site includes an 'in-frame' ATG initiation codon and *Xho*I sites are compatible with *Sal*I restriction sites which places the coding sequence 'in-frame' N-terminal to the *his* tag present within the vector.

Briefly, the pGAPZ expression vector was restricted with *Nco*I and *Sal* I and separated from the excised fragment by agarose gel electrophoresis; the plasmid fragment was excised and purified as above. Plamsid and insert fragments were then ligated.

Eschercshia coli TOP10 cells were then transformed with 3 µl of the ligation mixture with the pGAPZ construct containing the fragment of the required coding sequence. The colonies obtained were selected with 50 µg/ml of zeocin. Potential transformants were grown overnight and used to prepare plasmid DNA. The clones were then screened by restriction analysis of plasmid DNA, and selected clones were stored as glycerol cultures at -80 °C.

Using the pGAPZ clone candidate for protein expression in yeast, a loop of frozen cells was inoculated in 4 x 10 ml of LB zeocin (100 µg/ml) and incubated over-night at 37°C. Plasmid mini-preparations from these cultures were processed and pooled. The total DNA amount was estimated by analysis of 10 µl of the plasmid DNA by agarose gel electrophoresis, using known amounts of plasmid DNA as standards. 4 µg of plasmid DNA was linearized using 5 U of *Bln*I or *Rca*I. The correct enzyme to use was chosen according to the insert DNA sequence; to linearize the plasmid there must be no *Bln*I or *Rca*I site in the insert.

After ethanol precipitation and resuspension in 10µl of sterile water, the linearized DNA was added to 100 µl of chemically competent *Pichia* X33 using a sterile flow hood to avoid contamination. The transformation was carried out using the *Pichia* Easy Comp™ transformation Kit (Invitrogen™) according to the supplied protocol, with the modification that all volumes were increased two fold. Finally, transformed cells were selected by spreading the transformation mixture onto YPD agar plates, containing 100 µg/ml zeocin, and incubated for 3 days at 30°C.

Putative transformants were screened by colony PCR for clones containing the sequence to be expressed. Cells, individually picked from 6 yeast colonies using a sterile toothpick, were lysed by boiling for 5 min in 50 µl of PCR mix containing 10 mM dNTPs; 100 pmol/µl of each primer (pGAP FORWARD and 3'AOX); 120 mM Tris-HCl pH 8.5; 565.5 mM KCl; 25 mM $MgCl_2$. 1 U of *Taq* polymerase was added just before starting the PCR reaction. The PCR cycle contained a denaturation step of 1min at 94°C; a primer annealing step of 2 min at 50 °C; and a DNA extension step of 2 min at 72 °C, and was repeated 30 times. PCR products were analysed by a gel electrophoresis.

The clones that have given a PCR product of the correct size were used to produce medium-scale cultures. Culture medium was collected after 3 days culture in 250 ml of YPD and YPG medium at 30°C, by centrifugation at 4,000 xg for 20 min using a Beckman J2-HC centrifuge and a J2-21 rotor. The presence of recombinant proteins in the culture supernatant was determined by SDS PAGE, followed by Coomasie staining and Western Blot.

2.2.3.3.   *In vitro* coupled transcription-translation system

The TNT® Quick Coupled Transcription/Translation rabbit reticulocyte system (Promega) was used to produce protein directly from cloned coding sequences.

Bacteria containing the expression construct for AKIN11 were supplied by Dr. Victoria Lumbreras, Barcelona, Spain. Plamsid DNA was prepared from cultures of these bacteria.

50 ml of an over-night culture at 37 °C were processed in a DNA plasmid midi-preparation. Plasmid DNA prepared by this method was concentrated using a Speed-Vac centrifgual vacuum evaporator (Jouan, Ltd; Quarry Hill Road, Ilkeston Derbyshire, UK) to a concentration of at least 0.5 μg/μl. Concentration of the DNA was estimated by agarose gel electrophoresis as above. 2 μl of the prepared plasmid DNA sample was mixed with the components of the kit as described in the procedure provided. In 50 μl total volume, 40 μL of TNT$^{®}$Quick Master Mix containing reticulocyte lysate, reaction buffer, RNA polymerase, amino acid without methionine, and ribonuclease inhibitor were mixed with 2 μl of plasmid DNA (0.6 μg/μl), 6 μl of nuclease-free water and finally 2 μl of [^{35}S]-Methionine. The mix was incubated 1 h at 30 °C. To measure the incorporated radioactivity, proteins were TCA precipitated. 2 μl of transcription/translation products were treated with 98 μl NaOH 1 M incubated 10 min at 37 °C. The proteins were precipitated by addition of 900 μl of [25 % TCA 2 % casamino acids] and incubation on ice 30 min. 250 μl of the preparation was filtered using a Whatman$^{®}$ 3MM filter paper and a Whatman$^{®}$ GF/A glass fibre filter. The 3MM filter was immersed in liquid scintillation fluid and total incorporated counts determined using a scintillation counter.

2.2.4.   Protein Purification Procedures

2.2.4.1.   Glutathione affinity chromatography

GST-fusion proteins expressed in *E. coli* were purified using affinity chromatography on glutathione-agarose (Sigma$^{®}$) essentially as described by Sambrook and Russell, 2001.

E. coli containing GST-fusion constructs, in pGEX-5X.1, were grown overnight in 10 ml LB-broth with 100 μg/ml ampicillin at 37 °C. 12 hours later, each pre-culture was transferred into 200 ml LB-broth with the same concentration of ampicillin and grown until an OD_{600nm} of 0.4 was reached. Protein expression was induced adding IPTG to 1mM and the cultures were incubated for a further 4 H. Cells were harvested by centrifugation using a Beckman J2-HC centrifuge and a J2-21 rotor at 4,000 x g for 20 min.

Cells were lysed by sonication (4 x 1 min exposure, at 22 micron amplitude, using a Soniprep150 sonciator). After centrifugation at 10, 000 xg for 30 min, the resulting supernatant was half diluted in PBS buffer. All further manipulations were done at 4 °C. Glutathione-agarose beads were prepared by washing them in 10 volumes of distilled water twice and each time recovered by spinning 5 min at 1,000 x g. A last wash was made of 10 volumes of cold PBS. After spinning 5 min at 1,000 x g, the supernatant was discard and 1 volume of cold PBS was added to the beads. 2 ml of this slurry were added to the supernatant containing GST-fusion proteins derived from 100 ml of culture. The mix was shaken at 200 r.p.m. for 30 min at room temperature. Two washes in 10 volumes of PBS were carried out and an aliquot was kept for SDS PAGE analysis. Fusion proteins were eluted from the beads by addition of 1 volume of reduced glutathione 10 mM elution buffer. After shaking for 10 min at 200 r.p.m., the GST fusion proteins were recovered from the supernatant by centrifugation of the beads for 5 min at 1000 x g.

2.2.4.2.   Nickel affinity chromatography

$(his)_6$-tagged proteins were purified from protein mixtures in solution by affinity chromatography on nickel-agarose (Qiagen; Crawley, West Sussex, UK), using a small-scale batch procedure. Ni^{2+}-agarose beads were prepared by washing 3 times with excess loading buffer (5 mM imidazole), using a bench microcentrifuge (30 sec., 10,000 x g) to pellet the beads after each wash. The beads were then made up to a 1:1 slurry with loading buffer. They were then added to the protein extract, at

a ratio of 10 µl of slurry per 1ml of protein extract. After 30 min of incubation with rotation at room temperature the beads were washed 3 times with 1 ml of 5 mM imidazole, then with 1 ml 20 mM imidazole, and eluted with 1 ml of 300 mM imidazole.

2.2.4.3. Hydrophobic interaction chromatography

Proteins were purified from culture supernatants of *Pichia pastoris* by hydrophobic interaction chromatography. Proteins were precipitated by the addition of ammonium sulphate to 50 g per 100 ml solution, followed by incubation for 18 H at 4°C with stirring. The precipitated proteins were collected by centrifugation for 30 min at 3,000 xg. and resuspended in 1/100 of the initial volume with distilled water. A 2ml phenyl-Sepharose (Amersham Pharmacia) column (0.9 cm dia) was equilibrated with at least 5 column volumes of 4 M sodium chloride. The protein solution after ammonium sulphate precipitation was loaded on to the column and the column was washed with 4 M sodium chloride. A salt gradient [4 M → 0 M NaCl over 50 ml, followed by 20 ml water] was used to elute the column and 1 ml fractions were collected.

2.2.4.4. Assays: GST "pull down" assays

The GST "pull-down" assay is used to visualise protein-protein interactions where one partner in the interaction is a GST-fusion protein, which can interact with glutathione-agarose beads, and the other "target" protein is radiolabelled to allow its visualisation. The target protein was labelled by transcription-translation (section 2.2.3.3.). 25 µg of GST-fusion protein in 25 µl buffer (50 mM Tris-HCl pH 8.0) was incubated over-night at 4°C with 50 µl of [^{35}S]-methionine-labelled target protein and 50 µl of glutathione-agarose beads (prepared by making a 50% slurry in Tris-HCl buffer) with shaking at (150 r.p.m.). The mixture also contained protease inhibitor cocktail (Sigma®; 1 µl) and 66 µg/ml of antibiotics (ampicillin and kanamycin). The beads were separated by centrifugation in a microfuge (2 min at 12,000 x g) and the supernatant was removed and retained. The beads were then

washed 3 times with 1 ml buffer (containing protease inhibitor cocktail and antibiotics). Only the GST-fusion protein and the proteins interacting with it remained bound to the beads during these wash steps. Wash supernatants and the glutathione agarose pellets were retained for SDS PAGE analysis and autoradiography.

Chapter 3

3. Results

3.1. Identification and characterization of *At*PTPKIS1 by bioinformatics.

3.1.1.   The gene encoding *At*PTPKIS1

The occurrence of novel protein tyrosine phosphatases in plants was established by using a bioinformatics approach, exploiting the complete genome sequence of the model plant species, *A. thaliana*. Active protein tyrosine phosphatases (PTPs) contain a characteristic amino acid sequence motif, -HCXXXXXR-, where X is any amino acid (Fauman and Saper, 1996). A hydrophobic amino acid normally precedes the histidine residue in this sequence. This motif comprises the catalytically active cysteine residue in the enzyme, and two residues (histidine and arginine), which participate in catalysis.

An initial pattern matching search was carried out on predicted *Arabidopsis* proteins in the MIPS database (http://mips.gsf.de/proj/thal/db/search/search_frame.html) using the motif VHCXXXXXR. The initial valine was added to decrease the number of false hits, on the basis that it was present in the majority of protein phosphatases (Fauman and Saper, 1996). This initial search identified 27 predicted proteins, which are listed in Table 3.1. While a number of proteins were clearly false hits, on the basis of already having an assigned function, 7 of the proteins were predicted protein phosphatases, and a further 12 had no identified function. A second search was

Gene Designation	Predicted Protein	VHCXXXX XR motif	IPR000387 (PTP) Domain	Other Domains
At1g05000	Unknown protein	No	Yes	IPR001230 (prenyl group binding site)
At1g19110	Hypothetical protein	Yes	No	PF00092 (Von Willebrand factor type A)
At1g52780	Unknown protein	Yes	No	PF00426 (outer cpsid protein VP4)
At1g67120	Hypothetical protein	Yes	No	PF00158 (δ-54 interaction domain)
At1g71860	Protein tyrosine phosphatase	Yes	Yes	IPR000380 (procaryote DNA topoisomerase)
At1g72500	Hypothetical protein	Yes	No	IPR002035 (Von Willebrand factor type A)
At2g04550	Putative protein phosphatase	Yes	Yes	
At2g30640	Mutator-like transposase	Yes	No	IPR001878 (Zn-finger CCHC type)
At2g32960	Unknown protein	No	Yes	
At2g35680	Unknown protein	Yes	No	
At2g37130	Putative peroxidase ATP2a	Yes	No	PF00141 peroxidase
At3g02800	Unknown protein	No	Yes	
At3g06110	Putative dsPTP	Yes	Yes	
At3g09100	Putative mRNA capping enzyme, RNA guanylyltransferase	Yes	Yes	IPR001339 (mRNA capping enzyme), IPR000977 (ATP-dependent DNA ligase)
At3g10550	Putative myotubularin	Yes	Yes	IPR000510 (oxidoreductase, nitrogenase component 1)
At3g10940	Unknown protein	Yes	Yes	
At3g19420	Putative tyrosine phosphatase	Yes	Yes	
At3g23610	Dual-specificity protein phosphatase	Yes	Yes	
At3g50110	Putative tyrosine phosphatase	Yes	Yes	
*At3g52180	Putative protein	Yes	Yes	
At3g55270	Phosphatase – like protein	Yes	Yes	IPR001924 (gelsolin)
At4g19870	Putative protein	Yes	No	PF01344 (Keltch motif); PF00646 (F-box domain)
At4g22410	Putative protein	Yes	No	PF02148 (Zn finger C2H2 type)

Gene Designation	Predicted Protein	VHCXXXX XR motif	IPR000387 (PTP) Domain	Other Domains
At5g01290	mRNA capping enyme – like protein	Yes	Yes	PF001331 (mRNA capping enzyme)
At5g04540	Myotubullarin – like protein	Yes	Yes	
At5g16480	Putative protein	No	Yes	
At5g23720	Putative protein	Yes	Yes	IPR001680 (G-proteinbWD-40 repeats signature) IPB000130 (neutral Zn mettallo peptidase)
At5g28210	mRNA capping enyme – like protein	Yes	Yes	PF001339 (mRNA capping enzyme)
At5g38190	Putative protein	Yes	Yes	PF01637 (Archaeal ATPase)
At5g53490	Thylakoid luminal 17.4 kD protein, chloroplast precursor	Yes	Yes	
At5g56610	Putative protein	Yes	Yes	

Table 3.1. Predicted protein tyrosine phosphatases.

List of genes corresponding to the predicted protein, which includes or not the PTP signature motif and the domain IPR00387 characteristic of protein specific and dual-specific tyrosine phosphatase. The last column gives additional information about other domain recognized in protein sequence. Asterisk denotes the selected gene locus for this study.

carried out for proteins containing a domain characteristic of tyrosine specific- and dual specificity protein phosphatases, designated IPR000387. This search identified 20 predicted proteins, of which 18 had been identified by the initial search. The "hits" are listed in Table 3.1. Three sequences (*At*3g06110, *At*5g01290, *At*5g28210) encoding proteins with similarity to mRNA capping enzyme, which contained domain IPR001339 characteristic of this enzyme, were eliminated from further consideration. *At*1g71680, encoding a tyrosine-specific PTP, containing domain IPR000242, was also eliminated, as this gene has already been characterized (Xu *et al.*, 1998; Fordham-Skelton *et al.*, 1999). Further database searches were carried out with the full-predicted sequences of the remaining phosphatases and unknown proteins using BLAST software (both as the standard protein-protein similarity search BLASTP and the iterative matching similarity search PSI-BLAST; http://www.ncbi.nlm.nih.gov/blast/). Particular attention was paid to any protein shown to contain novel combinations of phosphatase and other domains, on the assumption that these proteins may be involved in plant-specific signal transduction processes.

From this analysis, a protein predicted by a gene locus designated *At*3g52180 (old designation F4F15.290) was selected for further study. The predicted protein is encoded by a gene on chromosome 3 of *Arabidopsis*, which occupies bases 19,362,552 – 19,358,688 (the gene is in reverse orientation compared to the numbering of this chromosome). 8 ESTs have been identified which correspond to this accession, which must therefore be a fully functional, expressed gene. The ESTs which were identified as corresponding to *At*3g52180 are listed in Table 3.2. Besides the dsPTP domain in the protein predicted by *At*3g52180, it also contained a C-terminal region which, on examination, had been incorrectly assigned due to intron-exon splicing not being correctly predicted by the database software. The predicted protein sequence could be corrected using sequence data from the ESTs; when this was done it became apparent that the protein contained a second domain. The possible functions of the predicted protein were explored further using BLAST searches, in which each predicted

EST	ORIGIN	TISSUE	SPLICING	NOTES
AV558174	cDNA clone SQ091c11F	Green siliques	ND	3' read
AV558705	cDNA clone SQ104e10F	Green siliques	Normal	3' read
AV562062	cDNA clone SQ163c03F	Green siliques	Normal	3' read
AV837117	cDNA clone RAFL09-85-C02	Mixed tissues	ND	5' read; complete ORF
BE528965	cDNA clone 600038162R1	Developing seed	Normal	5' read; complete ORF;
H36910	cDNA clone 180K17T7	Mixed tissues	Normal	5' read – ORF truncated at 5' end
T22081	cDNA clone 106E24T7	Mixed tissues	Alternative	5' read; complete ORFs
R84006	cDNA clone 106F24T7	Mixed tissues	Alternative	5' read; complete ORFs

Table 3.2.   ESTs corresponding to Arabidopsis thaliana gene At3g52180.

ND = not enough sequence data available to allow splicing pattern to be determined; ORF = open reading frame.

domain of the protein was compared to the global protein sequence database using PSI-BLAST (see above).

The similarity in the catalytic domain of the protein predicted by *At3g52180* extends over a region of approximately 140 amino acids, with greatest similarity to a class of PTPs from higher animals termed laforins (Minassian *et al.*, 1998; E value ≤ 5e -70 using the PSI-BLAST algorithm) and a PTP from *Chlamydomonas eugametos* (Haring *et al.*, 1995). Other PTPs had a lower level of similarity (E value ≤ 8e -31 using PSI-BLAST). Besides the VHCXXXXXR motif, the protein predicted by *At3g52180* also contained a conserved aspartic acid residue, diagnostic of an active phosphatase, which could function as a general acid during catalysis (Fauman and Saper, 1996). However, the predicted protein sequence does not contain a sequence motif characteristic of PTPs specific only for phosphotyrosine (–KNRY–), and thus the enzyme would be predicted to be a "dual-specificity" protein phosphatase (dsPTP), able to hydrolyse phosphoserine/threonine and phosphotyrosine residues. The most similar proteins identified by BLAST search are also dsPTPs. The protein predicted by *At3g52180* lacks a motif (-AY[L / I]M-) found in a subset of dsPTPs recently described in *Arabidopsis* (Gupta *et al.*, 1998; Ulm *et al.*, 2001), which includes MAP kinase phosphatases. Thus, on the basis of analysis of the sequence of the catalytic domain, *At3g52180* appeared to encode a novel dsPTP.

The second domain of the protein predicted by *At3g52180* was also compared to the global protein database using a BLAST search. In this case, sequence similarity was shown to the kinase interaction (KIS) domain of the β subunits of the AMP-activated / SNF1 / SnRK protein kinase complex in animals and yeast. E values ranging from 2e-44 to 4e-10 were obtained using the PSI-BLAST algorithm over the KIS domain region. The region of homology is located 23 residues C-terminal to the predicted phosphatase domain and extends over approximately 75 residues. The domain structure, and alignments of each of the

domains in the *At*3g52180 predicted protein with the corresponding most similar regions from other proteins are shown in figs. 3.1 and 3.2.

The protein predicted by *At*3g52180, from *A. thaliana* (*At*), contains both a protein phosphatase (PTP) and a kinase interaction sequence (KIS) domain, and was thus designated *At*PTPKIS1. It appears to be unique in the *Arabidopsis* genome. Further searching of the global protein databases suggested that homologues of this protein were present in other plant species (*q.v.*) but that no similar protein containing a combination of PTP and KIS domains existed outside the plant kingdom.

Plant proteins of this type may represent a novel type of protein phosphatase, which can interact with, and regulate, the SnRK protein kinase complex.

Homologues of *At*PTPKIS1 in other plant species

If *At*PTPKIS1 representes a new type of plant protein phosphatase with an important role in regulating metabolism, *via* the SnRK complex, then similar proteins would be expected to be present in other plant species. This was investigated by carrying out BLAST searches on the global protein database. Searching the non-redundant protein sequence database did not reveal any homologues of *At*PTPKIS1, which were similar in sequence over both domains. The search was therefore extended to the EST database. This search gave numerous ESTs, from a wide range of plant species which were similar to the PTP domain of *At*PTPKIS1, and a similarly large number with similarity to the KIS domain, although the similarities in this domain were generally less significant (according to e-score) than those found in the PTP domain. Three ESTs were identified which showed similarity to *At*PTPKIS1 over both the PTP and KIS domains; these are listed in Table 3.3. These ESTs come from three different plant families (Solanaceae, Compositae and Leguminosae), thus supporting the hypothesis that homologues of *At*PTPKIS1 are widespread in plants. The failure to find more sequences with homology over both domains reflects shortcomings

(a) PTP Domain


```
AtPTPKIS1    96-232
F9F8.24      91-227
F4P13.6     449-583
Q9WUA5      154-299
O95278      155-300
PTP3_CHLEU   95-206
```

```
AtPTPKIS1
F9F8.24
F4P13.6
Q9WUA5
O95278
PTP3_CHLEU
```

(b) KIS Domain


```
AtPTPKIS1     256-337
AtAKINβγ       22-101
ZmAKINβγ-1     26-105
F4P13.6       615-694
T7N9.13       454-531
AKIN_BETA1    102-179
AKIN_BETA2    103-180
potato_GAL83  105-182
AAKB_RAT       79-155
AAKC_HUMAN     78-155
GA83_YEAST    163-243
Q00995_KLULA  200-280
SIP2_YEAST    165-245
SIP1_YEAST    565-608
```

Fig. 3.1 Alignment of PTP and KIS domains in AtPTPKIS1 with corresponding domains in proteins identified as most similar in BLAST similarity searches.

(a) Comparison of the catalytic PTP domain of AtPTPKIS1, and related proteins F9F8.24 and F4P13.6, with laforin and *Chlamydomonas* PTP catalytic domains. The active site motif HCx5R is overlined and a conserved aspartate residue which could serve as a general acid during catalysis is shown with an asterisk. Accessions are mouse laforin [Q9WUA5], human laforin [O95278] and *Chlamydomonas* PTP3 [Q39491].

(b) Comparison of the KIS domain of AtPTPKIS1 with representative KIS domains in β subunits of AMP-activated, SNF1 and SnR kinases from animals, yeast and plants respectively. Also included are AtAKINβγ and ZmAKINβγ__, recently described SNF4 homologs which contain a KIS domain, accessions AC000106 and AF276085 repetively (Lumbreras *et al.*, 2001).

EST identifier	Species	Similarity score
AW040640	Tomato (*Lycopersicum esculentum*)	>200 in PTP domain, 80-200 in KIS domain
BG581666	Clover (*Medicago trunculatum*)	>200 in PTP domain, 80-200 in KIS domain
AW053458	Ice plant (*Mesembryanthemum crystallinium*)	>200 in a region encompassing most of PTP and KIS domains

Table 3.3. ESTs in other plant species with similarity to both PTP and KIS domains in AtPTPKIS1.

in the EST data; the nucleotide sequences which have been determined are generally not long enough to span both domains with sufficient accuracy to allow similarities to be detected.

The homologue from tomato was selected for further study; the cDNA was fully sequenced and was shown to encode a protein of 370 residues designated *Le*PTPKIS1. *Le*PTPKIS1 has a high level of similarity to *At*PTPKIS1 in both the phosphatase and KIS domains, leading to a predicted sequence identity of 63% overall for the two proteins. An alignment of the two coding sequences is shown in Figure 3.3. Work not reported in this thesis (Fordham-Skelton *et al.*, 2002) subsequently showed that *Le*PTPKIS1 was an active protein tyrosine phosphatase, thus confirming the functional activity assigned to the *Arabidopsis* homologue *At*PTPKIS1.

3.1.2.　　Characterization of *At*PTPKIS1 transcripts; alternate splicing

The eight ESTs identified which corresponded to *At*3g52180 have been listed in Table 3.2. Three cDNA clones (SQ104e10f, 180K17T7 and 106F24T7) from which the ESTs, AV558705, H36910 and R84006 respectively, are derived were fully sequenced. All three contained complete, or nearly complete transcripts, which could be used to confirm the sequence deduced for the mRNA and protein corresponding to the gene locus *At*3g52180, particularly with respect to intron-exon boundary assignments. SQ104e10f and 180K17T7 corresponded closely to the predicted intron-exon structure of the gene (although at least one splice site had been wrongly predicted by the analysis software); these cDNAs, and the mRNA they represent, are referred to as "full-length" *At*PTPKIS1 mRNA. cDNA 106F24T7, on the other hand, while unambiguously derived from the same gene, had a number of sequence differences when compared to cDNAs SQ104e10f and 180K17T7, which altered the intron-exon splicing pattern. This did not appear to be a sequencing artefact, as at least one other EST shows identical differences at its 5' end (see table 3.2). This cDNA, and the mRNA it represents, is referred to as "alternatively spliced" *At*PTPKIS1.

Fig. 3.2 Domain structure of protein KIS domain containing proteins derived from the genomic sequence data for *Arabidopsis thaliana*.


```
  1                                                                                351  TSD LI ER   AHA V  - LePTPKIS1, tomato
  1                                                                                359  LEE SK IQ   TCSPA V AtPTPKIS1, Arabidopsis
```

Fig. 3.3 Sequence comparison of AtPTPKIS1 with a tomato homologue, LePTPKIS1.

Active site motifs are highlighted. Identical residues and conserved substitutions are shown on black and grey backgrounds respectively. The sequence identity between the two predicted proteins is 63 %.

The alternatively spliced version of *At*PTPKIS1 has three differences at intron-exon splice sites when compared to the full length mRNA. First, there is an out-of-frame splice at the intron2-exon 3 boundary, deleting 4 bases from the mRNA. As a result of this aberrant splicing, a stop codon is introduced into the reading frame after amino acid residue 61. A second aberrant splice is present at the intron 5-exon 6 boundary, which introduces 28 extra bases into the "alternatively spliced" mRNA. Exons 7-11 are identical in each version of the mRNA, and this allows the reading frame to be established. Working back through exons 6 and 5 shows that the alternatively spliced mRNA has an open reading frame of 31 amino acid residues, differing from the corresponding residues in the full length protein, preceding residue 139 of the full length protein. The alternatively spliced mRNA also differs from the full length at the C-terminal end of the predicted coding sequence; intron 12 in the full length sequence is not spliced from the alternative form, resulting in 9 novel amino acid residues following exon 12 of the full length protein, and a stop codon. The alternatively spliced mRNA lacks the 43 amino acid residues of exon 13 of the full length mRNA. The alternatively spliced transcript therefore predicts a protein of 239 residues, truncated in both N- and C-terminal regions when compared to full-length *At*PTPKIS1 (figs. 3.4 and 3.5), and encodes a separate open reading frame corresponding to 52 residues at the N-terminal end of *At*PTPKIS1, with an extra 9 amino acid residues added. The 239 residues protein sequence modified by alternate splicing still contains the PTP active site motif and a complete KIS domain. However, the N-terminal deletion, accounting for 43 residues missing in the region of similarity to the PTP domain, could result in a non-functional phosphatase.

The identification of the alternatively spliced form of *At*PTPKIS1 in two separate ESTs in the database makes it highly unlikely that it is the result of a cloning or sequencing artefact. However, RT-PCR analyses using total RNA from mature leaf tissue of *Arabidopsis* could only detect the full-length form of *At*PTPKIS1 (data not shown), and none of the ESTs similar to *At*PTPKIS1

Fig. 3.4. Schematic diagram of the genomic region of AtPTPKIS1.

Alternate splicing produces two transcripts: above is represented the full length and below is represented the alternatively spliced.

Fig. 3.5. Deduced amino acid sequence, genomic organisation and alternate splicing of AtPTPKIS1

The nucleotide and deduced amino acid sequence of AtPTPKIS1 (top) and its alternatively spliced form (bottom) is shown. Introns are in lower case. Nucleotide numbering is from the ATG initiation codon. Differences in amino acid sequence in the alternatively spiced form are underlined and (---) indicates identical sequence. The active site motif 'HC5xR' is underlined in bold. Arrows indicate differences in splice site selection: black for the full length form and grey for the alternatively spliced form.

(1-70) GGGACTGAGAAAGTAGCAGCTTCGTTTGCTTCCTCTGATATTTTTCTACACTCTCTGTCTCTCTCTAAAG

ATGAATTGTCTTCAGAATCTTCCCAGgtatctttcttgctttttcgaaattagcaaatgttgttgatttgtgtggttctttg 80
M N C L Q N L P R
- - - - - - - - -

attcgatatgaatttgtgtgttgttgatgtctttgatcgtcatttattcagATGTTCAGTCTCACCTCTGCTGGGATTC 160
                                                  C S V S P L L G F

GGGTGCATTCAAAGAGATCATTCTTCTTCTTCTTCTTCTTTGAAGATGCTAATATCGCCTCCGATCAAAGCCAATGATCC 240
G C I Q R D H S S S E S D K M L I S R P I K A N D P

AAAAATCTCGACTTGTTTTACATgtatacaccctctttaagatatattgttatatcagtttttttctctaactttttattg 320
K S R L V L H
- - - - - - -

acacttttgtaagtgctttgagtgtagGCAGTATCAGAGTCAAAATCCAGCTCAGATGAGTGGTGTTGCAAAGGATGA 400
        gcagTAT
        A V S E S K S S S E M S G V A K D E
        Y Q S Q N P A Q E *

AGAGAAATCTGATGAATATAGCCAAGACATGACTCAAGCTATGGGTGCTGgtaaattgttttcatcatacctatacatgt 480
E R S D E Y S Q D M T Q A M G A

ccttagattggaaaactctaatgatgtccaccattttttgtctcatattggaaatattgatttagctcattggattgtatt 560
ctactcttgtttttcatcttcatcttgattagtttttgttcatctgtctaccaaggaatcttatatcttttgcttgctta 640
ttgacaaacttatcttcgtttctccttaattatgtatttcagtcactgattcatctcatgtgtctttcatttcagTTC 720
                                                                              V

TAACTTACAGGCACGAGTTAGGAATGAACTACAACTTTATTCGTCCAGATCTAATTGTTGGATCCTGCTTACAGgttcac 800
L T Y R H E L G M N Y N F I R D P L I V G S C L Q

ttttatatccttctctcaaatgatgattttgttttggatatgtggtacttcctcgtgttcttactgtatgctctcctt 880

attgctagACCCCTGAAGATGTTGACAAGCTTCGTAAAATTGGAGTTAAAACCATATTTTGCTTGCAACAAGATCCAGAC 960
        T P E D V D K L R K I G V K T I F C L Q Q D P D
        M L T S F V K L E L K P Y F A C N K I Q T

CTGGAgtatccttccaaagccctctttttctgttcatgattatgattttgttttcactttccttcgcattgagtgaggaa 1040
L E
W

gaatgatctgcagtgcagaatgccttgaccaatccttccagATATTTTGGAGTAGACATAAGCAGCATCCAAGCCTATGC 1120
-------------------------TGCAGAATGCCTTGACCAATCCTTCCAG---------
        M Q N A L T N P S R
        Y F G V D I S S I Q A Y A

TAAGAAATATAGTGATATTCAGCATATTCGCTGTGAAATTAGgtaaggcaacgatctagctgtgacttttttaatatattg 1200
K K Y S D I Q H I R C E I R

gtattcaaagctgatgcgtaatgaatcgaaggagagatcatgagttgattttttagtttccttccaatatttttcttatt 1280
gcttgacttttataagaagcatatgaggtagtttttttttccagaattatctgaataggcgagggaacaaaggtatatacac 1360
tgttggagggtagtctcagatatctttaggcacgcaatctttaaattcttattgttttgtcagtgaagtgttatactttt 1440
cttttttctttgggctctttacttttatcacgtttcctaagctaatgggaaatcagcacacttttgtgtaagcaataactg 1520
gtttatgacgaattgtgagtgatagtttttttttttaccgggtgtgaaatcactgctatcaactaactttttttgttcgtcca 1600
ccttTgtctgtttacaaactagagtaactgtagaagttagtcatcatcccattctcactaaatttctttactgctaaaga 1680
cttggaaccatcatcccttttggcatatatgagactgttcacgagcacatgcatagaaaaattcaccatggacgtac 1760
ccactcacaacgtaggggattgacatattatatatcattgatgtaacagAGACTTTGATGCATTTGATTTGAGAATGCGT 1840
                                                  D F D A F D L R M R
                                                  - - - - - - - - - -

CTTCCAGCGGTGGTTGGTACTCTTTACAAAGCTGTTAAGCGAAATGGAGGAGTTACATATGTGCACTGCACTGCTGGAAT 1920
L P A V V G T L Y K A V K R N G G V T Y V H C T A G M

GGGAAGGGCTCCTGCTGTTGCGgtatgttacagaaacctgttactaagctcttctctcttttttcccccttttctttccagta 2000
G R A P A V A

ttgaaaaaatgtaagactgttgcaaacagtagttgtcttaattttgtgcttatttggttgactcatgaaggtatgaacaa 2080

gtttttcacattttgctgactgatactgagaagggatattaattcatttcagTTGACATACATGTTCTGGGTGCAAGGC 2160
                                                    L T Y M F W V Q G

TATAAGCTTATGGAAGCTCATAAATTACTTATGgtaagatactaatttacttgtgtggtttatgttctctatatgtgttttct 2240
Y K L M E A H K L L M

ttatacttatattaagtaaccaacatgttaatcatgtttttactagtgttactgoctgttatattatgtgaggatgtttgg 2320
ggatttccttcttctcctatgaaccgacaattatgataactagaatcatagaccattgggtcagggggtttgtagtaaata 2400
tacactgaagatatagcacttttaaaagacaaaaatcttgggttagttcatccaaaatagtctagcattagacattttbag 2480
tttgatatctctagatggctggtgaaacataaagcacgcgtaaatttggtttatataacatttctgaaatttttcttgca 2560
gAGCAAAAGGTCGTGCTTTCCGAAGCTGGATGCTATCAGAAATGCAACAATTGATATTgtgagtcagcgttcaactttca 2640
S E K S C F P K L D A I R N A T I D I
- - - - - - - -

caatgcttcctatgtttgaaaaacctcagttccattgattagcatgagtgcgcaacagttacctggttccttgactcctg 2720
attattgtttccgggtgtatatggataaaagCTTACAGGACTCAAGAGGAAGACTGTTACTCTGACACTGAAAGATAAGG 2800
                                L T G L K R K T V T L T L K D K

GGTTCTCCAGAGTAGAAATTTCTGGCCTTGACATTGGATGGGGACAGgtaaatatattttatcttaaaccaatttattta 2880
G F S R V E I S G L D I G W G Q
- - - - - - - - -

ttatgacaccttgctcttaaatctatggtgtgccatataagctttaaagccaggatgcacttttgtcactgctattaaaa 2960
gaatggcctctttgataagaagtggaagagatgagtgttggatttctgaaattattgtcttatgattgtgcaaaactat 3040
ttggacgtatatatataaaaggtaatgtatttctttagctcttttcgttttttaccattttctctctttatgtagAGGATAC 3120
                                                                            R I
                                                                            - -

CTCTAACACTGGACAAGGGAACAGGATTCTGGATCCTAAAGAGAGAACTGCCTgtgagttttcttcttaactgttaagca 3200
P L T L D K G T G F W I L K R E L P

tagtgtctgatcattttccaacaaccactacaatttctgaaatgcagGAAGGACAGTTTGAATATAAATACATCATAGAT 3280
                                                E G Q F E Y K Y I I D

GGTGAATGGACACACAATGAGGCCGAACCGTTTATAGGACCTAACAAAGACGGCCATACCAACAATTACGCTAAAgtaaa 3360
                                                                              *
                                                  ---------------------------GTAAA
G E W T H N E A E P F I G P N K D G H T N N Y A K
                                                                              - V K

aatgtctctgtctcttaacatatagctgaacaagtttttgatgaatgagttaatgctgtaatgttgttgtgtactaatgaa 3440
AATGTCTCTGTGTCTCTTAACATATAGCTGAACAAGTTTTGATGAATGAGTTAATGCTGTAATGTTGTTGTGTACTAATGAA
M S L S L N I *

ttagGTAGTGGACGATCCAACAAGTGTGGATGGTACAACTCGGGAGAGACTATCGAGCGAAGACCCTGAGCTGTTGGAGG 3520
TTAG--
V V D D P T S V D G T T R E R L S S E D P E L L E

RAGAACGCTCGAAACTAATCCAGTTCTTGGAGACTTGTTCTGAGGCAGAAGTTTGAGAGTTGAGAGCATCCGACTAGACT 3600
E E R S K L I Q F L E T C S E A E V *

ATGTGAATGTACAACTTACTAGTAACAATAAAATGCTTCATGTTAGTAATACATACAGTTACGCTTATAGCAGCAAAAGA 3680

AAAAAACATGAAGTTAGTAGTACTATGTTCAAGTATAATATATAGTTTAAAAETTTGTTTTTTCTTAATT (3751)

identified in other plant species (table 3.3) show an alternative splicing pattern, all resembling the full length mRNA over the regions sequenced. Further characterization of the occurrence, and the functionality (if any) of this alternately spliced form of the *At*PTPKIS1 mRNA will need to be carried out before any conclusions can be drawn about its possible role in the plant.

3.2. Interaction between the KIS domain in *At*PTPKIS1 and the SNF1-related kinase AKIN11.

3.2.1. Introduction.

Although bioinformatic approach described in section 3.1 suggested that the novel protein phosphatase *At*PTPKIS1 contained a KIS domain, which would interact with SNF1-like protein kinases, this hypothesis must be confirmed by experimental demonstration of the interaction. Assay systems are available which allow the interaction to be tested either *in vivo*, by using a yeast two hybrid assay system, or *in vitro*, by direct measurement of complex formation. In order to carry out such assays, a suitable "target" to which *At*PTPKIS1 would be expected to bind was needed. *Arabidopsis* contains 3 genes encoding SNF1-related protein kinases, one of which is a pseudogene. The two SNF1-related kinases are termed AKIN10 and AKIN11 (Bhalerao *et al.*, 1999). AKIN11 has been shown to interact with other KIS domain-containing proteins (Lumbreras *et al.*, 2001). On this basis AKIN11 was selected as the "target" for *At*PTPKIS1 binding.

3.2.2. Binding of *At*PTPKIS1 to AKIN11 *in vivo* in yeast two-hybrid assays.

Interaction between the KIS domain of *At*PTPKIS1 and the *A. thaliana* SNF1-like kinase AKIN11 *in vivo* was investigated by using yeast two-hybrid system (Fordham-Skelton *et al.*, 2002).

Coding sequences corresponding to full length *At*PTPKIS1, as well as the alternatively spliced *At*PTPKIS1 and the KIS domain of *At*PTPKIS1 alone,

were expressed in fusion with the yeast GAL4 activation domain *via* expression constructs prepared in the plasmid pACT2 (fig. 3.6.). This plasmid has been specially designed for the yeast two-hybrid system. pACT2 contains the coding sequence of GAL4 activation domain situated 5' to its cloning site region, so that when pACT2 is transformed into yeast cells, expression of the recombinant gene results in fusion proteins, which contain the GAL4 activation domain N-terminal to the inserted coding sequence (Durfee *et al.*, 1993). Constructs were prepared by using PCR with appropriate primers (see methods section) to amplify coding sequences corresponding to full length *At*PTPKIS1, alternatively spliced *At*PTPKIS1 and the KIS domain (residues 256-337 of the full length sequence), using the full length cDNA clones characterized in section 3.1 as templates. The PCR primers incorporated an *Nco*I and *Xho*I site. PCR products were purified by gel electrophoresis and cloned into an intermediate vector (pCR2.1); clones were characterized by restriction digestion. Inserts were excised from the intermediate vector by digestion with *Nco*I and *Xho*I, then cloned in the plasmid pACT2 which had been restricted using the same enzymes. The resulting recombinant plasmids were purified from selected clones, and checked by restriction and partial DNA sequencing. A similar construct was assembled to allow AKIN11 to be expressed in fusion with the GAL4 DNA-binding domain (fig. 3.6.) utilising the vector pAS2 (Harper *et al.*, 1993). These constructs were combined in *Saccharomyces cerevisieae* YRG2 by co-transformation.

The yeast two-hybrid system utilises a yeast strain which contains two modified genes with gal4 promoters, but which lacks the endogenous transcriptional activator protein, which would normally allow transcription to occur (Field and Song, 1989). The protein consists of two domains, a DNA-binding domain, and the transcription activation domain; both domains are required for activity. The GAL4 DNA-binding domain binds to the yeast DNA 5' to the gal4 promoter, and positions the activation domain in such a way that pACT2 can initiate the formation of a transcription complex at the promoter (fig. 3.6). If the two domains are separated, no transcription is possible unless they can

Fig. 3.6. Yeast two-hybrid system.

Diagram explaining the mechanism of yeast two hybrid system Where AD represent the GAL4 activation domain as BD= GAL4 DNA binding domain fusion; *HIS3* reporter gene situated in the yeast strain genome. Reporter gene *HIS3* is only expressed if AD can be brought into proximity with BD; in this case the binding domain allows the activation domain to interact with the promoter. Protein domains to be tested are fused to AD and BD, and brought together, expression occurs.

be made to interact to form a complex. Thus, if sequences fused to the GAL4 binding and activation domains interact, and bring the two domains together, expression of the genes occurs. Yeast strain YRG2 contains a selectable indicator gene, *his*3, which confers histidine auxotrophy, and a screenable indicator gene, *lacZ*, which gives a blue phenotype to colonies grown in the presence of a chromogenic galactoside substrate (X-gal).

Yeast colonies were transformed with the two-hybrid system plasmids pAS2 and pACT2 containing AKIN11/GAL4-BD (binding domain) and *At*PTPKIS1/GAL4-AD (activation domain) fusions respectively, and various controls. Colonies containing both plasmids were selected. Screening of transformed colonies, for interaction between the GAL4 domains, was initially carried out by plating transformed yeast on histidine-deficient medium. The pAS2 vector without any insert was used as a negative control; yeast transformed with this vector plus pACT2 construct showed no growth on selective media. Interaction between AKIN11 and the *At*PTPKIS1 protein would permit the gal4 promoter to be activated and the indicator genes to be expressed. The resulting gene products allowed yeast to produce histidine to offset the lack of histidine in the medium, and thus grow. All three fusions, with full length *At*PTPKIS1, alternatively spliced *At*PTPKIS1 and the *At*PTPKIS1 KIS domain gave positive results (fig. 3.7.) in the assays, and allowed yeast to grow on histidine-deficient media, showing that interaction of *At*PTPKIS1 with SnRK occurred, and was mediated by the KIS domain. To confirm the result, expression of the reporter gene *lacZ* was assayed in colony filter lift assays. However, these assays failed to give a positive blue phenotype, even after a longer incubation with X-gal at 37 °C. This observation could be due to the host used, which was a derivative of HF7c. This genomic background gives strong activation of *his*3 indicator gene but relatively weak activation of the *lacZ* reporter gene. Thus a second assay was carried out in which the two-hybrid plasmids were transformed into yeast strain Y190, which has been shown to give stronger activation of the *lacZ* reporter gene. In this strain, the KIS domain – GAL4-AD fusion gave a strongly coloured blue

Fig. 3.7. Interaction of AtPTPKIS1 with AKIN11 in the yeast two-hybrid system shown by activation of *His3* reporter gene.

(a) represent active protein fusion construct used for each assay.
(b) Assay in two different media where lack of leucine and tryptophane are selective reagents of the yeast strain used; first three plate: medium with lack of histidine, selection by *HIS3* reporter activate; negative control on half of each plate without the AKIN11 construct; second three plate: medium where histidine is available.

(a)

(b)

phenotype in the colony lift assay in the presence of X-gal (fig. 3.8.), whereas the other fusions gave a weakly coloured blue phenotype. These results are in agreement with those observed with the activation of the *his3* indicator gene, and confirm the interaction between *At*PTPKIS1 and AKIN11, mediated by the *At*PTPKIS1 KIS domain, in this system.

3.2.3. Binding of *At*PTPKIS1 to AKIN11 *in vitro* in GST pull-down assays.

3.2.3.1. Preparation of GST-KIS fusion protein.

The results given by the yeast two hybrid system assay, described above, demonstrate interactions between AKIN11, with the full length and alternatively spliced *At*PTPKIS1, and with the KIS domain on its own. To confirm these results, and prove that they are not an effect of mediation by endogenous yeast proteins, an assay of the interaction *in vitro* was necessary (fig. 3.9. details the different step for the preparation of the biding test).

Assembly of GST-KIS expression construct.

To confirm the interaction between the KIS domain of *At*PTPKIS1 and AKIN11 by an assay *in vitro*, a GST-KIS domain fusion protein was produced by expression in *E. coli*. The expression construct encoded a recombinant fusion protein containing a glutathione S-transferase (GST) sequence fused N-terminally to the *At*PTPKIS1 sequence, and was prepared using the vector pGEX.5X.1. This vector contains the coding sequence of the GST from *Schistosoma japonicum* (Smith and Johnson, 1988). The GST gene fusion system allows the expression of GST fusion proteins (in *E. coli*) that can be purified by affinity chromatography using glutathione-agarose beads. The GST gene also allows fusion proteins to be detected using colourimetric or immunological methods, like western blotting with an anti-GST antibody (fig. 3.10.). Some methods use the strategy of affinity chromatography of GST fusions to demonstrate protein: DNA or protein:protein interactions (GST pull down).

Fig. 3.8. Interaction of AtPTPKIS1 with AKIN11 in the yeast two-hybrid system shown by activation of *lacZ* reporter gene.

FL = AtPTPKIS1, "Full length".
ALT = AtPTPKIS1, alternatively spliced.
KIS = AtPTPKIS1, KIS domain only.
GAD = GAL4 activation domain.
GBD = GAL4 binding domain.

Fig. 3.9. Preparation of the GST Pull Down assay.

Fig. 3.10. Construct for expression of GST-KIS domain fusion protein.

The assembly of the expression construct is shown in the upper panel; the lower panel shows the expected fusion protein product produced by transcription and translation.

To assemble the expression construct, the KIS domain from *At*PTPKIS1 (residues 256-337 from the full-length protein sequence) was amplified by PCR from the appropriate cDNA (SQ104e10f), using PCR primers incorporating *Eco*RI and *Xho*I sites as described in the methods section. The PCR amplification gave the desired product as a band of 400bp (see fig. 3.11.). The band containing the required DNA fragment was excised from the gel, and the DNA was purified, ligated to the intermediate cloning vector pCR2.1, and cloned in *E.coli*. Plasmid DNA was purified from colonies picked at random, and tested for the presence of the desired insert by restriction digestion with *Eco*RI and *Xho*I. Because *Eco*RI and *Xho*I sites were present in the PCR primers, the expected result was two bands, representing the plasmid and the insert respectively, at approx. 3.5 Kb and 400 bp. Clones containing an insert of the correct size were selected; one clone was chosen for checking by sequencing of the plasmid DNA, which confirmed that the correct fragment had been cloned and that no sequence errors due to PCR were present.

The insert was then excised from the intermediate vector by restriction with *Eco*RI and *Xho*I, purified by agarose gel electrophoresis, and ligated with pGEX5X.1, which had also been restricted with *Eco*RI and *Xho*I, and purified by agarose gel electrophoresis. The ligation mixture was transformed into competent *E. coli* cells, and plated out on selective media. Potential recombinant clones were screened by preparation of plasmid DNA, followed by restriction digestion with *Eco*RI and *Xho*I. Clones giving the expected band pattern, a vector band of 4.9 kbp and an insert band of 400 bp, were obtained (fig. 3.12.). A suitable clone was selected, and plasmid DNA from the clone was subjected to DNA sequencing to ensure that the correct fusion had been produced. This clone was then used for the expression studies.

Expression and purification of GST-KIS protein.

The clone containing the validated GST-KIS domain fusion expression construct was grown in liquid culture, and expression of the recombinant protein was induced by adding IPTG. A clone containing the vector only (which should

Fig. 3.11. Amplification of the KIS domain

DNA electrophoresis analysis of the KIS domain after PCR amplification.
M = 1Kb DNA size marker. KIS domain DNA band is indicated by arrow.

Fig. 3.12. Restriction diagnostic of KIS/pGEX.5X.1.

DNA electrophoresis separating pGEX.5X.1 from the KIS domain after *Eco* RI and *Xho* I restriction digest. M = size marker Lambda DNA/*Eco*471 cut with *Ava* II; bands at 4900 bp correspond to the vector and faint bands at 400 bp correspond to the KIS domain. Construct was verified by sequencing.

produce GST) was used as a positive control for expression. Cells were harvested from 1 ml samples of the cultures before and after induction, and proteins were analysed by SDS-PAGE. Results are shown in fig. 3.13. Both the GST-KIS fusion construct and the control showed polypeptides that were present as abundant components of the total proteins after induction, but were not present before induction. The GST positive control had an induced polypeptide band at approx. 26 kDa, the correct molecular weight for GST, whereas the GST-KIS fusion construct had an induced polypeptide band at approx. 35 kDa, corresponding well with the expected size of the fusion protein (26kDa GST + 10kDa KIS domain).

The GST-KIS fusion protein was purified from *E. coli* total cell proteins by affinity chromatography on glutathione-agarose, utilising a batch method (Sambrook and Russell, 2001). The clone-containing vector only was again used as a positive control. Samples of the purified proteins were analysed by SDS-PAGE; results are shown in fig. 3.14. Both GST and GST-KIS fusion proteins were obtained free from contaminants by this method; no other proteins bound to the agarose-glutathione beads after processing. Whereas the purified GST gave a single band on gel, the GST-KIS fusion gave two closely spaced bands, with some evidence for a small amount of free GST also being present. This suggests that the KIS domain on the fusion is susceptible to proteolytic degradation in the *E. coli* cells, and that approx. 30% of it had lost a few amino acids from the C-terminus, but only a very small fraction had undergone significant degradation. Degradation must be occurring in the KIS domain, since the GST itself is stable. The gel analysis of purified proteins by SDS-PAGE shown in fig. 3.14. contains tracks which were loaded with different amounts of samples, to evaluate the concentration of each protein by densitometry of the stained gel. The purified GST was used as a standard, since its concentration could be independently determined by absorption at 280 nm, using a literature value for the absorption coefficient. Results are shown in fig. 3.15. The preparation of GST-KIS fusion protein was estimated as being almost exactly four times less concentrated than the

Fig. 3.13. Expression of GST and GST-KIS proteins in *E. coli*.

SDS-PAGE analysis of total cell proteins from *E. coli* transformed with the expression constructs encoding GST or GST-KIS fusion protein. Tracks UI = uninduced cultures, I = induced cultures, M = molecular weight markers. Induced polypeptides are indicated by arrows. Expression of GST show a band at 26kDa, Expression of GST-KIS at 35 kDa not visible in the uninduced cell lysate.

Fig. 3.14. Purification of GST and GST-KIS.

SDS PAGE analysis of the GST and GST-KIS protein fusion purified by glutathione agarose. M = molecular weight marker. Different amounts of samples were loaded to evaluate the concentration.

GST preparation, and this corresponded to concentrations of 3.22 µg/µl for GST and 0.86µg/µl for the GST-KIS fusion.

3.2.3.2. Preparation of labelled AKIN11 by coupled transcription-translation *in vitro*.

A radioactively labelled AKIN11 target for use in the GST pull down assay was prepared by an *in vitro* coupled transcription/translation reaction, using a commercial system (TNT® Quick Coupled Transcription/Translation rabbit reticulocyte system; Promega). In this system, a rabbit reticulocyte lysate (Pelham, and Jackson, 1976) was used to translate RNA transcribed from a recombinant plasmid (based on pET28a) containing an AKIN11 coding sequence under the control of a T7 promoter. By adding T7 RNA polymerase to purified plasmid DNA in the presence of ribonucleotides and amino acids, AKIN11 polypeptides are produced.

The clone of AKIN11 in pET28a was kindly supplied by Lumbreras (2001). Plasmid DNA was prepared by a standard minipreparation protocol, and adjusted to a concentration of 0.6 µg/µl. The template DNA was then added to a transcription-translation reaction according to the protocol supplied with the reagent kit. [^{35}S]-methionine was added as radioactive label for the synthesised polypeptide. A negative control with water and a positive control with luciferase T7 control DNA, provided with the kit, were subjected to an *in vitro* transcription-translation under the same conditions. Protein synthesis was monitored by incorporation of radioactivity into TCA-precipitable material. Results are shown in table 3.4. Incorporation of radioactivity in the AKIN11 reaction was approx. 80% that in the positive control, with background incorporation less than 10% of total, and thus the synthesis of labelled AKIN11 was successful.

3.2.3.3. GST pull-down assay.

The purified GST-KIS domain fusion protein was incubated with the labelled AKIN11 polypeptide produced by transcription-translation *in vitro* (see above).

	Count (CPM)	Count minus background (CPM)
−	294	
+	3879	3585
AKIN11*	3147	2853

Table.3.4. AKIN11 insertion of [35S]-methionine.

Scintillation counts analysis after protein TCA precipitation of in vitro transcription translation products. - = negative control made of water, [^{35}S]-methionine plus reagent from the kit; + = positive control made of luciferase T7 control DNA, [^{35}S]-methionine plus reagent of the kit; AKIN11* = protein to label made of AKIN11/pET28a plasmid contruct, [^{35}S]- methionine plus reagent of the kit; CPM = counts per min.

Fig. 3.15. Purification of GST and GST-KIS

Comparison of GST and GST-KIS protein fusion concentration observed on SDS PAGE analysis. V = volume loaded on the gel, n = quantity estimated of the samples, n_1 = quantity estimated by densitometry of the stained gel of GST-KIS, n_2 = known quantity of GST. The first diagram represents the estimated concentration of both proteins: GST=3.22µg/µl; GST-KIS=0.86µg/µl. The second diagram shows GST-KIS as four time less concentrated as GST.

Glutathione-agarose beads were then added to the incubation, in excess, to bind all free GST-KIS fusion and any GST-KIS:AKIN11 complex. The beads were pelleted by centrifugation, washed, and then treated with SDS-sample buffer to dissociate any proteins from the matrix. Released proteins were then analysed by SDS-PAGE, followed by autoradiography. If the labelled AKIN11 has formed a complex with GST-KIS, then a polypeptide band of the correct size for AKIN11 should be visualised after autoradiography, resulting from the labelled AKIN11 in the GST-KIS:AKIN11 complex bound to the beads. Two negative controls were processed in the same conditions as the assay with GST-KIS plus AKIN11. The first control contained labelled AKIN11 only with glutathione-agarose beads. The second contained labelled AKIN11 with GST (not the fusion) and the glutathione-agarose beads. Results of the GST pull-down assay are shown in fig. 3.16. Both negative controls gave the expected result of no band after autoradiography, showing that AKIN11 does not bind to the glutathione-agarose matrix on its own, nor does it interact with GST. However, the reaction between AKIN11 and the GST-KIS fusion protein gave a band of the correct size for AKIN11 after autoradiography, showing that interaction between the two components had taken place *in vitro*. This result confirms that the KIS domain in *At*PTPKIS1 is able to interact with the SNF1-like kinase AKIN11, as indicated by the yeast two-hybrid system.

3.3. Expression of Functional *At*PTPKIS1 in *Pichia pastoris*.

*At*PTPKIS1 is predicted to be a dual specificity protein phosphatase on the basis of its sequence similarities to proteins of known function. To demonstrate enzymic activity and confirm the identification of *At*PTPKIS1 as a dsPTP, by showing that it hydrolyses substrates containing phosphotyrosine, phosphoserine and phosphothreonine residues, functional recombinant protein is required. Attempts to produce functional *At*PTPKIS1 using a pET expression system in *E. coli* BL21 were unsuccessful, insofar as the protein produced was insoluble, and could not be refolded to produce a functional enzyme. In contrast, the

Fig. 3.16. GST pull down assay: protein interaction GST-KIS *versus* AKIN11.

Autoradiography analysis of the GST pull down assay. AKIN11* = AKIN11 protein probe with [^{35}S]-methionine light the film by its presence in the supernatant and in the pellet corresponding to GST-KIS that immobilized it *via* their interaction; - = negative control made of Tris-HCl pH 8.0 buffer; GST = negative control made of GST in Tris-HCl pH 8.0 buffer.

*At*PTPKIS1 homologue from tomato, *Le*PTPKIS1, could be refolded after expression in *E. coli*, and gave a functional enzyme, which was active as a phosphatase (Fordham-Skelton, 2002). To confirm the functional activity of *At*PTPKIS1, expression of the full length *At*PTPKIS1 has been undertaken using a yeast expression system to produce folded, recombinant protein, which should show functional activity.

3.3.1. *At*PTPKIS1 cloning in pGAP expression vector.

The expression host chosen was *P. pastoris*, allowing the use of an integrating plasmid system for transformation. The choice of the vector was the pGAP plasmid, containing the yeast gap promoter derived from the gene encoding glyceraldehyde phosphate dehydrogenase. The gap promoter directs constitutive high level expression in *P. pastoris* (Waterham *et al.*, 1997). The vector also directs secretion of recombinant proteins into the culture medium, determined by the presence of sequences encoding the native *Saccharomyces cerevisiae* α–factor prepro-sequence in frame with the inserted coding sequence (Cregg *et al.*, 1993). A poly-histidine tag can be engineered on the C-terminal of the recombinant protein by cloning in-frame with the vector sequence, which allows immuno-enzymatic detection using anti-His antibody or Ni^{2+}-probe, and affinity purification on a column containing immobilised Ni^{2+}. Details of the steps taken to perform this yeast expression of *At*PTPKIS1 can be found fig. 3.17.

The vector pGAPZαA was used as a basis for the construct to express *At*PTPKIS1. This vector contains a replication origin allowing it to be manipulated in *E. coli* as a standard cloning plasmid, but functions as an integrating vector in *Pichia*, with a selectable antibiotic resistance gene (zeocin resistance). The vector was modified by replacement of the *Eco*RI site in the multiple cloning region with an *Nco*I site, and removal of an *Nco*I site elsewhere in the vector, to leave a unique *Nco*I site in the multiple cloning region. The *At*PTPKIS1 insert was excised from the construct prepared for the yeast two-hybrid system (see above) by digestion with *Nco*I and *Xho*I, purified by extraction from a gel slice after

FL-AtPTPKIS1cDNA

pACT2
(in Top 10)

$\begin{bmatrix} NcoI \\ EcoRI \end{bmatrix}$

pGAPZαA

FL-AtPTPKIS1

pGAPZαA

(in Top 10)

•Restriction diagnostic
•sequencing

Pichia pastoris X33
transformed with the
linearized plasmid

•Screening by colony PCR
•Expression of the selected clone

•Ammunium sulfate
precipitation
•HIC purification
•Ni^{2+}agarose purification

•Ammonium sulfate precipitation
•Dialysis
•Ni^{2+}agarose purification under
denature condition 8M urea

Fig. 3.17. Production of functional AtPTPKIS using *Pichia pastoris*.

agarose gel electrophoresis, and ligated to the modified pGAPZαA vector that had also been restricted with *Nco*I and *Xho*I. The ligation mixture was transformed into *E. coli* competent cells, and transformants were selected by plating on selective media. Screening of the clones was performed by preparing plasmid DNA from four colonies, band analysis by restriction digest with *Bam*HI. This enzyme is predicted to cut the construct at three sites: one just before the cloning site, one inside the insert, and one at the end of the insert sequence. Results are shown in fig. 3.18. Three bands corresponding to the plasmid at 3.9 kbp, and the insert fragments at 620 bp and 590 bp, were observed as predicted. Addition of the two fragments confirms the correct size of the full length *At*PTPKIS1 DNA (1.15 kbp). The sequence of one of the clones was checked by DNA sequencing to ensure that no errors were present. DNA was prepared from this validated clone of the expression construct for *At*PTPKIS1 in pGAPZαA, ready to be transferred in *P. pastoris*.

3.3.2.   Production of clones of *P. pastoris* expressing *At*PTPKIS1.

3.3.2.1.   Transformation and preliminary screening of the *At*PTPKIS1 clones.

4 μg of DNA of the validated expression clone (see last section) were linearized by restriction with *Bln* I. (This restriction enzyme cuts the pGAPZαA plasmid at one site that is distant from the cloning site; *At*PTPKIS1 has no sites for this enzyme in its full length coding sequence.) After ethanol precipitation and resuspension in sterile water the linearized plasmid was used to transform *P. pastoris* X33. The transformed cells were plated on selective media containing zeocin. After 2 days, colonies were visible. Six colonies were tested for the presence of the introduced DNA by colony PCR, using two specific primers which hybridize to the extremities of the pGAP cloning region (5'pGAP:468→487; 3'pGAP:902→925). The amplification products were analysed by an electrophoresis gel (fig. 3.19.). All clones show a band at 1.15 Kb on the gel, which corresponds to the expected size of the *At*PTPKIS1 construct. This result shows that these are genuine

Fig. 3.18. Restriction diagnostic of AtPTPKIS1/pGAPZαA

DNA Electrophoresis separating pGAPZαA from AtPTPKIS1 after BamHI restriction digest. M = size marker Lambda DNA/Eco471. Band at 3900 bp correspond to the vector and faint bands at 620 bp and 590 correspond to AtPTPKIS1.

Fig. 3.19. Yeast colony PCR of AtPTPKIS1 in *Pichia pastoris*.

DNA Electrophoresis on 1 % TAE-Agarose separating PCR product from other polynucleotides. M = size marker: lambda DNA/Eco471. Dinstinct bands at 1300 bp are visible for each clones analysed with a better amplification for the first two clones.

transformants, and that the selection system has worked correctly, but gives no indication of expression of the recombinant polypeptide. Screening of the clones 1 to 6 by colony PCR had to be followed by screening for the expression level of the encoded polypeptide before choosing a clone for expression on a larger scale.

To screen for *At*PTPKIS1 protein expression, small scale batch cultures (50ml) of the yeast clones were set up, with *P. pastoris* containing pGAPZαA without insert as a negative control. As a result of using the yeast expression system, the proteins were released in the culture supernatant, and were soluble. Culture supernatant was separated from cells by centrifugation, and proteins were separated by acetone precipitation. The precipitated proteins were resuspended in SDS loading buffer, and analysed by SDS-PAGE. Coomassie blue staining of the SDS-PAGE gel was not efficient enough to detect *At*PTPKIS1 mixed with *P. pastoris* proteins, but a Western blot probed with an HRP-conjugated nickel reagent (Pierce) was able to detect a putative product. As shown in the fig. 3.20., the Western blot shows a single band for full length *At*PTPKIS1 in clones 1 to 4, a faint band for clone 5, and none for clone 6 or the negative control (*P. pastoris* pGAPZαA). Clone 1, which produced more *At*PTPKIS1 protein than the other clones, was selected to be grown in a larger scale batch culture.

3.3.2.2. Production of *At*PTPKIS1 in *P. pastoris*.

Clone 1 was grown as a 1 litre batch culture. Proteins were separated from the culture supernatant by ammonium sulphate precipitation. A portion of the precipitate was analysed by SDS-PAGE and Western Blotting, (fig. 3.21.) using the same detection system (Pierce) as before. The Western blot did not give a clear result, as the salt in the ammonium sulphate pellet interfered with the gel electrophoresis, but a positive reaction was obtained with the detection reagent, and three bands corresponding to the protein were obtained. This might indicate partial hydrolysis of the *At*PTPKIS1 under the large scale culture conditions. A negative control culture made with *P. pastoris* X33 with pGAPZαA vector only, was grown at the same time and in the same culture conditions, and analysed by

Fig. 3.20. Screening of six clones presumed *At*PTPKIS1/pGAPZαA in *Pichia pastoris*.

Autoradiography of a western blot using chemiluminescence reaction to a specific conjugate His probe HRP. Analysis of six clones positives on colony PCR. The proteins were culture supernatant extracted and precipitated by 80% acetone. + = positive control, known His tag protein 30kDa; - = proteins extract from *Pichia pastoris* transformed by pGAPZαA without insert.

Fig. 3.21. Western blot of *At*PTPKIS1 after ammonium precipitation.

Autoradiography of a western blot using chemiluminescence reaction to a specific conjugate His probe HRP. Analysis of the full length *At*PTPKIS1 His tag after ammonium precipitation of dialysis of proteins from a selected clone showing a band after analysis of the proteins acetone precipitated. + = positive control, known His tag protein 30kDa; - = proteins extract from *Pichia pastoris* transformed by pGAPZαA without insert.

the same methods. No bands are visible for the control protein precipitate, which means that no unspecific proteins coming from contamination or endogenous yeast proteins bind to the nickel. A positive control made of a known *His*-tag protein proves that the test functioned correctly. The predicted size of *At*PTPKIS1 is approx. 41kDa. The bands obtained from clone 1 are difficult to size accurately, but span a size range of approx. 35kDa – 60kDa.

Proteins present in the ammonium sulphate precipitate from culture supernatant of clone 1 were separated using hydrophobic interaction chromatography (HIC) on a column of phenyl-Sepharose. Proteins were applied in 4M NaCl and eluted using a NaCl gradient: from 4M to 0M. Proteins were collected as 85 fractions, and selected fractions were analysed by Western Blotting (fig 3.22.). Two peaks of protein were eluted from the HIC column. The first peak, fractions 58-61, contained a polypeptide of approx. 60kDa, corresponding to the upper band of the three bands present in the loaded material, whereas the second peak, fractions 73-79, contained two polypeptides, an upper, stronger band of approx. 45kDa, and a lower, weaker band of approx. 35 kDa; these two bands appear to correspond to the lower two bands in the starting material. Fractions corresponding to the two protein peaks were pooled separately.

The two protein peaks from the HIC column were separately further purified by affinity chromatography on a nickel column. However, no bands appeared when the wash and eluted fractions were analysed on a Western Blot. The *At*PTPKIS1 protein, despite the presence of a *His*-tag, did not appear to bind to the nickel resin. To check for the possibility that the poly histidine tag might be folded inside the protein, and be inaccessible to bind to the affinity matrix, purification under denaturising conditions, in 8 M urea, was carried out. For this experiment, a new batch culture was prepared, from which the culture supernatant was not pre-purified by HIC. Instead, the culture was dialysed against 5 mM imidazol, and applied directly to the column. Although Western blotting analysis showed that bands detected by the nickel reagent were present in the culture

Hydrophobic Interaction Chromatography

Fig. 3.22. Hydrophobic interaction chromatography of *At*PTPKIS1.

Upper: Elution profile of column.
Lower; Western blot of column fractions probed with a His-specific HRP conjugate.
 + = positive control, known His tag protein 30kDa.
Arrows indicate bands from His-tagged polypeptides, putatively *At*PTPKIS1.

supernatant (fig. 3.23.), no binding to the nickel column, even in the presence of 8 M urea, was observed.

3.4. Investigations of potential function of *At*PTPKIS1

3.4.1. Analysis of gene expression using microarray data

Microarray technology, where a large number of individually immobilised target sequences representing a subset of the expressed genes (both known and predicted) of an organism are hybridized to a labelled mRNA preparation, has permitted an important expansion in the study of gene expression (Schena *et al.*, 1999). By assessing the strength of signal from the hybridization reaction, a measure of the presence and semi-quantitative amount of a specific mRNA in a mixed population can be determined. The occurrence of mRNAs encoded by specific genes in particular tissues, or under particular conditions, can then lead to hypotheses for the function of the encoding gene, where this is not known, based on similarity of the expression pattern to those of other characterized genes. For example, if an unknown gene is upregulated under conditions of heat shock, in the same way that known heat shock genes are up-regulated, it is reasonable to assume that the unknown gene has a role in protection of the tissue against the effects of this stress condition (Kurian *et al.*, 1996). In this way, the functions of a number of genes (both known or unknown) have been identified and organized in computer databases (Ermolaeva et al., 1998). Although expression of the gene *At*PTPKIS1 has not been specifically studied by this microarray approach, data for its expression proved to be available in public databases, since some of the ESTs corresponding to the gene had been used as target sequences in microarrays used for gene expression studies. Information was available for the ESTs 106E24T7, 106F24T7 and 180K17T7 (see Table 3.2).

Data were taken from the work of Schaffer *et al.* (2001) who analysed the expression of a large number of *A. thaliana* genes using a microarray system, to define which genes were regulated by the diurnal and circadian rhythms during plant growth. Microarray data were found on two websites: http://genome-

Fig. 3.23. HisTag *At*PTPKIS1 purification

Autoradiography analysis of a western blot using chemiluminescence reaction to a specific conjugate His probe HRP. Analysis of the His Tag *At*PTPKIS1 batch purification under denaturated conditions (UREA 8M). + = His Tag protein known as 30kDa; S = supernatant; $W_1 \rightarrow W_4$ = Wash number 1 → wash number 4; E = Elution samples. Bands corresponding to *At*PTPKIS1 are visible in the supernatant and not in the eluted sample.

www5.stanford.edu.edu/ and http://www.prl.msu.edu/circl. Analysis of data from the three ESTs of *At*PTPKIS1, one corresponding to the full-length (108K17T7) and two (106E24T7 and 106F24T7) corresponding to the alternatively spliced mRNAs (Fig. 3.24.), revealed that *At*PTPKIS1 gene expression varies with the circadian and diurnal rhythm. *At*PTPKIS1 expression was up-regulated during the second half of the light period (*i.e.* the afternoon) in plants exposed to a 12 H dark:12 H light regime (Fig. 3.25.). Results from all three ESTs were similar, showing that the conclusion was not the result of an artefact. Expression of *At*PTPKIS1 was also upregulated by dark, and was higher in roots and dark-grown tissues, whereas in light-grown tissues, such as leaves, its expression was lower. Cluster analysis applied to the whole dataset from these experiments reached similar conclusions about *At*PTPKIS1 expression (see Schaffer *et al.*, 2001).

Expression of the gene coding for the subunit AKINβ1 of the SnRK AKIN11 has been studied as well, using the same dataset. AKINβ1 contains the KIS domain, which in the SnRK-like complex allows the activation of α subunit. Two ESTs, 212O21T7 and 314D1T7, corresponded to this gene, and allowed its expression pattern to be compared to that of *At*PTPKIS1. Results (Fig. 3.24) showed that the circadian rhythm also influenced AKINβ1 expression, but when compared to *At*PTPKIS1, the effect on expression level was opposite; when *At*PTPKIS1 was up-regulated, AKINβ1 was down-regulated and *vice-versa*. For example, AKINβ1 is more expressed in the morning than in the afternoon, contrary to *At*PTPKIS1, which is upregulated in the afternoon.

The variation of the expression of the gene encoding *At*PTPKIS1 with circadian rhythm, which is opposite to the expression of the SnRK subunit AKINβ1, strongly suggests that, like AKINβ1, *At*PTPKIS1 plays a role in regulating the SnRK, but that its role is opposite to that of AKINβ1.

Fig. 3. 24. *AtPTPKIS1* and *AKINβ1* show opposite patterns of diurnal expression.

Significantly increased transcript abundance is denoted by red squares and decreased transcript abundance by green squares between the time indicated above the cluster followed or not by a criteria that defines the type of experiment and repetition: dark, continuous dark regime; light, continuous light regime; biological, repetition that used different plants ; single, experiment that was not repeated ; no criteria, light/dark cycle of 12 hours each repeated using the same mRNA. Data were taken from the work of Schaffer *et al.* (2001) describing experiments designed to identify genes where expression are controlled by circadian rhythms. Cluster analysis indicated that AtPTPKIS1 is upregulated in the afternoon whilst that of AKIN1 is upregulated in the morning. At least two independent ESTs corresponding to each gene were analysed. EST clone identifiers were for AtPTPKIS1; 106E24T7, 106F24T7, 180K17T7 and for AKINβ1; 212O21T7, 314D1T7. Data is available at http://genome-www5.stanford.edu/ and http://www.prl.msu.edu/circl.

Fig. 3.25. Reciprocal regulation of *AtPTPKIS1* and *AKINβ* expression

Red corresponds to significantly increased transcript abundance and green to significantly decreasing transcript abundance between the type of tissue or treatment, or between the time indicated above the cluster followed or not by a criteria that defines the type of experiment and repetition: dark, continuous dark regime ; light, continuous light regime ; biological, repetition that used different plants ; single, experiment that was not repeated ; no criteria, light/dark cycle of 12 hours each repeated using the same mRNA. Arrows denote the major differences in expression patterns between the two genes. Clone Ids for ESTs corresponding to *AtPTPKIS1* and AKINβ1 are shown. Data was taken from publicly available microarray websites (http://genome-www5.stanford.edu/) and the microarray website corresponding to the microarray analysis of *Arabidopsis* genes regulated diurnal and circadian rhythms (Schaeffer et al., 2001; see cluster analysis at http://www.prl.msu.edu/circl/816Aft.gif).

3.4.2.   *Arabidopsis* mutants containing gene knockouts

Data presented in this thesis lead to the conclusion that the novel protein *At*PTPKIS1 regulates metabolism by interaction with the SnRK complex in *A. thaliana*. The availability of plants containing a putative knockout of the *At*PTPKIS1 gene gave the possibility that a phenotype resulting from the absence of *At*PTPKIS1 might be observable in planta. Simultaneously, knockouts for other putative dual specificity protein phosphatases were observed to attempt to identify phenotypes resulting from the absence of the corresponding proteins.

The publicly available collection of *A. thaliana* (Columbia) knockout lines (Syngenta) was used to provide lines containing knockouts in a series of genes; the lines were provided as seed from the transformed plants. Six plants of each line were grown. The knockout lines contained insertions in the following genes: *At*1g27070 (not yet identified but presumed to be part of the SnRK complex as it contains domains with high similarity to the AMPK/SNF1 βγ subunits); *At*3g52180 (*At*PTPKIS1), *At*3g10940 (putative dual specificity protein phosphatase), *At*3g01510 (putative dual specificity protein phosphatase), *At*1g78540 (putative dual specificity protein phosphatase) and *At*2g35680 (putative dual specificity protein phosphatase containing transmembrane segments). No apparent phenotype was noticeable on the mutant lines, either when grown on artificial media in culture, or when transferred to soil, apart from a weak growth for a few plants (apparently at random), which could have been due to poor culture conditions.

Chapter 4

4. Discussion

This project aimed to identify novel protein tyrosine phosphatases in plants through the use of bioinformatics databases, and then to use a molecular biological and biochemical approach to investigate their potential function *in vivo*. It has focussed on a predicted protein containing both a dual-specificity protein phosphatase domain, and a domain with similarity to the *K*inase *I*nteraction *S*equence present in components of the SNF1 kinase complex in other organisms. The protein has thus been designated *At*PTPKIS1. The presence of the KIS domain suggests that this protein is able to interact with, and putatively regulate the activity of, the plant SNF1 kinase complex; since the SNF1 kinase itself functions as a global metabolic regulator, *At*PTPKIS1 itself is also predicted to have a role in metabolic regulation.

The bioinformatics approach aiming to identify putative proteins was based on the complete genome sequence for *A. thaliana*, which became available in 2000. The availability of this dataset had been thought to allow the proteins present in an organism to be fully defined, but it has become apparent that predicting the complete set of proteins in an organism (the "proteome") from the genome sequence is not entirely straightforward. A number of factors combine to increase the diversity of proteins when compared to that of genes. First, proteins can undergo post-translational modification, by glycosylation, addition of lipids, modification of amino acid residues, and proteolytic cleavage, and these post-translational modifications can be tissue-dependent. Secondly, alternative splicing

of transcripts from genes can result in multiple mRNA molecules being produced by a single transcription unit, and thus multiple proteins being produced. Thirdly, products of single genes can combine to form proteins containing multiple subunits, and the composition of these protein complexes can vary with tissue type, development or external stimuli. As well as these factors, which are not inherent in the genomic sequence data, shortcomings in the analysis of the DNA sequence data can make it difficult to predict protein sequences. Some protein sequences may be missed altogether when genomic sequence data is analysed, because the coding regions (open reading frames) are not recognised; this is particularly likely to be the case for very short coding regions. Secondly, for many genes, intron-exon boundaries have been identified by automated software algorithms, which are not completely accurate, making the predicted protein sequences incorrect. The "discovery" of *At*PTPKIS1 described in this thesis was dependent on the correction of incorrect assignments of intron-exon boundaries made by software algorithms, by comparison of sequences from cDNA clones (ESTs) with the genomic DNA sequence. The EST sequences that are available are generally too short to allow this operation to be carried out over the whole length of a gene, and therefore it is likely that the *Arabidopsis* genomic database still contains many incorrectly predicted protein sequences. It will be necessary to check predicted sequences for processed transcripts against full-length cDNA sequences, until all the predicted protein sequences have been properly confirmed.

Dissection of bioinformatic databases for *A. thaliana* to find novel PTPs was based on two criteria. Proteins containing the motif -HCX$_5$R- corresponding to the catalytic domain found in all PTPs were obtained using MIPS database (http://mips.gsf.de/proj/thal/db/search/search_frame.html). 27 proteins were recovered, 12 were hypothetical or unknown proteins, and 7 were predicted PTPs, three of which had already green characterized: one tyrosine-specific PTP, *At*PTP1 (Xu *et al.*, 1998) and two dual specificity PTPs *At*dsPTP1 (Gupta *et al.*, 1998) and *At*MKP1 (Ulm *et al.*, 2001). The three proteins correspond to the genes *At*1g71860, *At*3g23610, *At*3g55270 (table 3.1.). The second database search was

based on a sequence pattern characteristic for PTP/dsPTPs, designated IPR00387. 18 proteins contain this pattern, 3 which are not protein tyrosine phosphatases (do not contain PTP motif) and 8 were unknown or predicted dsPTPs. Using overlapping search criteria in this way gives a higher degree of confidence to predictions of function than using a single search criterion. Several of the predicted proteins have interesting sequence features; for example, the protein predicted by *At*2g35680 (unknown protein, contains PTP motif) is actually a predicted trans-membrane dual specific PTP (http://plantsp.sdsc.edu/cgi-bin/detail.cgi?at_number=At2g35680/). This would be the only trans-membrane protein phosphatase existing in plants. Use of the PSI-BLAST algorithm (http://www.ncbi.nlm.nih.gov/blast/) to identify a putative dual specific PTP (corresponding to the gene *At*3g52180) containing a domain with high similarity (E-value = $2e^{-44}$) to the KIS domain of the β subunit of AMP-activated kinase, SNFI and SnRK, which plays an important role in sugar sensing and signalling has been described earlier. *At*PTPKIS1 has a catalytic domain similar to other dual specific PTPs, but lacks the motif -AY[L/I]M- characteristic of dsPTPs known to function as MAP kinase phosphatases. *At*dsPTP1, which has been actually identified as a MAP kinase regulator (Gupta *et al.*, 1998) contains this motif. *At*PTPKIS1 is unique as a fully characterized gene that encodes a protein containing a dsPTP domain and a KIS domain fused together. No other protein sequence predicted by genomic analysis of any organism shows this feature. This protein is thus of great interest, in that it may be part of a metabolic regulatory mechanism unique to the plant kingdom.

*At*PTPKIS1 represents a class of protein found generally in higher plants, since it has homologues in other plant species. Although homologues have only been found in dicots including tomato (*Lycopersicum esculentum*), clover (*Medicago trunculatum*) and ice plant (*Mesembryanthemum crystallinium*), it seems likely that similar proteins will also be identified in monocots as the sequence databases are enhanced.

Full sequence data from selected ESTs (table 3.2.) corresponding to the gene *AtPTPKIS1* also identified the possibility of alternate splicing of transcripts of this gene. The alternately spliced product is of a smaller size (239 amino acids against 347 for the full length) due to truncations of the N- and C-termini. 43 amino acid residues of exon 13 in the full-length product are replaced by 9 that follow the exon 12 at the C-terminus of the alternately spliced product. The N-terminal end is lacking the first 55 amino acid residues, with 34 residues of the 3rd exon replaced by 9 novel amino acids. Much more importantly, the active site region is modified by omission of the 4th exon and by a different translation of exon 5. These alterations may change the functional properties of the protein. However, the catalytic motif and the KIS domain are preserved. The alternately spliced transcript has a short open reading frame, with start codon, 5' to the reading frame similar to full length *AtPTPKIS1*, and thus to produce the alternately spliced protein would require that translation was initiated at an internal ATG codon. The sequences around the potential initiation codons do not indicate which one might be preferred, but not using the most 5' ATG for translation initiation is not usual. The possibility must be considered that the alternatively spliced transcript is an artefact of the cDNA synthesis reactions, possibly a fusion of two independent cDNA sequences. Evidence against this is that two ESTs with the alternative splicing pattern have been identified, which appear to differ in length by a few bases at the 5' end but are otherwise identical, and that the sequence similarity between the alternately spliced and full length transcripts over the whole of their extent makes a fusion impossible. However, the alternately spliced form could not be detected by RT-PCR of RNA from leaf tissue, and no similar alternately spliced EST has been found in any other species. Characterization and functional properties determination are needed to validate the alternately spliced transcript as a component of the plant cell. In this thesis, results obtained in the yeast two-hybrid system show that the alternately spliced protein retains functionality in the KIS domain at least, which provides further evidence that alternative splicing in this gene is not an artefact.

The kinase interaction sequence (KIS) domain plays a vital role in maintaining the integrity of the AMP-activated kinase (AMPK) multi-protein complex. Since the AMPK complex has a functional role in nutritional stress regulation (Hardie *et al.*, 1998), proteins containing functional KIS domains, which can interact with the kinase component of the complex, are also potential metabolic regulators. As described earlier, the "normal" AMPK complex contains three subunits: α, the kinase subunit, which is activated by phosphorylation and inactivated by dephosphorylation; β, a "scaffold" subunit; and γ, a regulatory subunit. The β subunit contains the two protein binding domains, KIS (interacts with α) and ASC (interacts with γ) that maintain the integrity of the complex (see fig. 4.X). While plants contain genes which encode proteins similar to all three of the components of the "normal" AMPK complex, the occurrence of other predicted proteins containing domains which could interact with components of the AMPK complex (see Fig. 4.1.) suggests that modulation of the activity of this global metabolic regulator may be more complex than in yeast or animals. The identification of *At*PTPKIS1 introduces the possibility of a protein phosphatase which can specifically dephosphorylate the plant homologue of the kinase subunit of the AMPK complex, thus inactivating it, by direct interaction *via* a KIS domain. For *At*PTPKIS1 to be able to play this role, both the protein phosphatase domain and the KIS domain must be functionally active. The aim of this thesis has been to demonstrate that this is the case.

Demonstration of the protein phosphatase activity of *At*PTPKIS1 unfortunately was not achieved. Both the full length and alternatively spliced versions of the polypeptide were expressed in *E. coli*, but the resulting protein was found exclusively in the insoluble fraction of total proteins extracted from cells (see Fordham-Skelton *et al.*, 2002), even when expression was carried out at reduced temperature (30°C instead of 37°C). The expressed recombinant proteins could be purified under denaturing conditions using a C-terminal his-tag added to the expression construct, but attempts to refold the protein by removal of the

denaturing agent were unsuccessful (data not presented). In contrast, the tomato homologue of *At*PTPKIS1, *Le*PTPKIS1, was partially soluble when expressed in *E. coli* at 30°C from a similar construct, and could be purified under non-denaturing conditions to yield a functional protein. *Le*PTPKIS1 hydrolysed the generic phosphatase substrate, pNPP (p-nitrophenyl phosphate), although the phosphatase activity expressed on a protein basis was low when compared to the tyrosine specific protein phosphatase *At*PTP1 (Xu *et al.*, 1998; Fordham-Skelton *et al.*, 1999). This low activity against pNPP is observed in other dual-specificity protein phosphatases, e.g. *At*dsPTP1 (Gupta *et al.*, 1998). *Le*PTPKIS1 activity was inhibited by vanadate and phenyl arsine oxide, inhibitors specific for dsPTPs (Fordham-Skelton *et al.*, 2002). The protein also dephosphorylated two phosphotyrosine-containing peptides, and thus has been demonstrated to be a functional protein phosphatase. On the basis of sequence similarity, it is reasonable to assume *At*PTPKIS1 has a similar function.

It is not clear why *Le*PTPKIS1 can be successfully expressed in *E. coli* whereas *At*PTPKIS1 cannot. The two proteins are very similar in sequence except in the N-terminal region, and this seems the most likely cause of the failure of *At*PTPKIS1 to fold in *E. coli*. A construct where the sequence of *At*PTPKIS1 was truncated at the N-terminus might give a product which could be folded successfully. Interestingly, when the sequences of both *At*PTPKIS1 and *Le*PTPKIS1 are analysed for predicted cellular location using the PSORT programme (psort.nibb.ac.jp; Nakai *et al.*, 1992), *At*PTPKIS1 is predicted to be located in the chloroplast, probably in the stroma, whereas *Le*PTPKIS1 is predicted to be cytoplasmic. This would suggest that the N-terminal region of *At*PTPKIS1 may be a chloroplast targeting sequence. A similar result for *At*PTPKIS1 is given by the TargetP (v 1.1) subcellular location prediction software (www.cbs.dtu.dk; Emanuelsson *et al.*, 2000), although using this prediction algorithm, *Le*PTPKIS1 is predicted to be mitochondrial in location. The sequences were also analysed by the chloroplast transit peptide prediction software ChloroP 1.1 (www.cbs.dtu.dk; Emanuelsson *et al.*, 1999.); using this algorithm both

proteins were predicted to contain N-terminal chloroplast transit peptides, that of *At*PTPKIS1 being 54 residues, and that of *Le*PTPKIS1 45 residues. Given the overall similarity of the two proteins in sequence, and the fact that *Arabidopsis* has only a single gene of this type, it seems unlikely that their subcellular locations differ, so that if one protein is located in the chloroplast, the other will be also. Although the predictions are not wholly consistent or certain, it is certainly possible that *At*PTPKIS1 contains an N-terminal chloroplast targeting sequence. If this is the case, removal of the appropriate N-terminal region of the protein would facilitate folding in a prokaryotic environment. The shorter predicted chloroplast targeting sequence of *Le*PTPKIS1 may allow it to fold in *E. coli* without the sequence being removed.

As has been found for other eukaryotic proteins which could not be expressed in *E. coli* in a soluble form, the use of *P. pastoris* as an expression host for *At*PTPKIS1 yielded a soluble product, secreted into the culture medium as directed by the N-terminal fusion with the α-factor preprosequence contained in the expression vector. However, this system gave a very low yield of product, and it was not possible to purify or characterize the recombinant *At*PTPKIS1. Although the product could be detected on Western blots using a reagent specific for the C-terminal his-tag in the expression construct, affinity purification of the recombinant protein on a nickel-agarose column, or by batch processing with nickel-agarose resin, was unsuccessful. These results were inconclusive in demonstrating that the *Pichia* expression system could produce *At*PTPKIS1 as a functional phosphatase, as the protein could not be assayed. Despite all the evidence supporting the conclusion that *At*PTPKIS1 is an active protein phosphatase, a direct demonstration that this is the case remains to be achieved.

In contrast to the difficulties encountered when trying to demonstrate the function of the phosphatase domain in *At*PTPKIS1, the function of the KIS domain in this protein could be readily established. The KIS domain of *At*PTPKIS1 can be expressed in *E. coli* as a soluble polypeptide (Fordham-Skelton

et al., unpublished data), which can be purified under non-denaturing conditions. Therefore, it is not surprising that the GST-KIS domain construct produced for the GST pull-down assay and expressed in *E. coli* also gave a soluble product. This assay was used to confirm results obtained in the yeast two-hybrid system, where all three proteins, *At*PTPKIS1, the alternatively spliced version of *At*PTPKIS1 and the KIS domain gave a positive result for interaction with an *Arabidopsis* SNF1-like kinase (see section 3.3). The functional homology of the KIS domain in *At*PTPKIS1 to other KIS-domain containing proteins, inferred from sequence similarity, has thus been confirmed by direct assays.

However, although the KIS domain in *At*PTPKIS1 clearly is able to interact with SNF1-like kinases in *Arabidopsis*, whether it actually does so *in planta* has not been established. In order for such an interaction to take place, both the kinase and *At*PTPKIS1 would need to be expressed in similar tissues at similar developmental stages, and be present in the same subcellular compartment. If, as predicted from sequence analysis (see above) *At*PTPKIS1 is located in the chloroplast, then it will not be available to interact with cytoplasmic cellular components. Further, in order for *At*PTPKIS1 to form a stable complex with the kinase, the dissociation constant for the complex under cellular conditions would need to be sufficiently low. The two-hybrid and pull-down assays suggest that the interaction between *At*PTPKIS1 and AKIN11 is relatively weak when compared to positive controls; activation of the reporter or selectable gene in the two-hybrid system is partial, and the pull-down assay does not show precipitation of all the available labelled AKIN11. In addition, the two-hybrid system suggests that the interaction gets weaker in the sequence: KIS domain > alternatively spliced *At*PTPKIS1 > *At*PTPKIS1, although this may reflect the relative levels of expression of the proteins in yeast rather than the inherent strength of any interaction. Even if *At*PTPKIS1 does not form a stable complex with AKIN11, the interaction *via* the KIS domain may still serve to direct this protein phosphatase to a specific target, and dephosphorylate it. There need not be a stable complex formed for the dephosphorylation to function in regulation of

Fig. 4.1. Model representative of different association found in the SNF1 complex.

i) Yeast SNF1 like complex.
ii) AKIN11 like complex.
iii) AKINα/ AtPTPKIS1 interaction representation.

metabolism; in addition, a stable complex with *At*PTPKIS1 could prove disadvantageous when the kinase needs to be rephosphorylated.

The simple model for *At*PTPKIS1 function, interaction with and dephosphorylation of the SNF1-like kinase (α-subunit of the SnRK complex) in *Arabidopsis*, would give this protein an important role in metabolic regulation. The SNF1 kinase is activated by phosphorylation, and can be inactivated by dephosphorylation by a protein phosphatase (Ludin *et al.*, 1998). In dephosphorylating the α kinase subunit of an SnRK complex, *At*PTPKIS1 might displace it from the other parts of the complex *via* an interaction mediated by the KIS domain on the phosphatase. However, in considering potential roles for *At*PTPKIS1, interactions other than those within the trimeric structure of the SnRK complex need to be taken into account. It is clear that the basic heterotrimeric structure does not account for all the interactions, which can take place with components of the SnRK complex. For example, the activated SNF1 kinase in yeast is dephosphorylated by a heterodimeric protein phosphatase complex (Ludin *et al.*, 1998), which contains a PP1 serine / threonine phosphatase subunit, and a specific binding subunit (Reg1) which interacts with the SNF1 kinase subunit at a site distant from the interactions with other subunits of the SNF1 complex. In addition, other types of protein can interact with plant SnRKs. In *Arabidopsis* these include PRL1, a nuclear regulatory WD-repeat containing protein, which interacts with both the C-terminal regions of the *Arabidopsis* SnRKs AKIN10 and AKIN11 (Bhalerao *et al.*, 1999), a nuclear import receptor (Nemeth *et al.*, 1998), and an ubiquitin ligase subunit, SKP1/ASK1 (Farras *et al.*, 2001). Therefore, given the emerging complexity of SnRK interactions, it is possible that the "target" for *At*PTPKIS1 dephosphorylation is not a component of the canonical complex, but another protein (or proteins) associated within the SnRK complex. The protein phosphatases which dephosphorylate activated α kinase subunits of SNF1 and AMPK complexes in yeast and animals are serine/threonine specific, and belong to a functionally different class of enzyme from *At*PTPKIS1 (Hardie *et al.*, 1998).

The last aim of the project was to investigate potential functions for the novel protein *At*PTPKIS1 in the plant cell. The major part of this work was based on analysis of the expression of the encoding gene, using publicly available microarray data from the two web sites http//genome-www5.stanford.edu.edu/ and http://www.prl.msu.edu/circl. These sites contain analyses of data for hybridization of *Arabidopsis* mRNA extracted from different tissues, and under different environmental conditions, to microarrays containing a wide variety of ESTs, including those for *At*PTPKIS1.

AKINβ1 is the regulatory β subunit of the *A. thaliana* SnRK-like complex. In plants, the β subunit contains only the KIS domain, but not the ASC domain, although the β subunit still functions as in yeast through its association to the α subunit, which allows activation of the kinase by phosphorylation (Lumbreras *et al.*, 2001). Regulation of expression of the genes *AtPTPKIS1* and *AKINβ1* (Fig. 3.24.) shows a remarkably clear pattern, with the two genes being oppositely affected by the circadian rhythm. *AtPTPKIS1* is up-regulated in non-photosynthetic tissues in the afternoon when *AKINβ1* is up-regulated in the morning. These data provide evidence for *At*PTPKIS1 and AKINβ1 functioning as mutually antagonistic regulators of the SnRK complex kinase activity, through phosphorylation mediated by AKINβ1 and dephosphorylation mediated by *At*PTPKIS1.

The SNF1 family is known as metabolic regulators. In mammalian cells, AMP-activated protein kinase inactivates the 3-HMG CoA reductase (sterol and isoprenoid synthesis) by phosphorylation in rat liver (Beg *et al.*, 1973) and is regulated by fructose in rat hepatocytes (Gillespie *et al.*, 1992). The major role of the yeast SNF1 is the activation of glucose repressor genes (Gancedo *et al.*, 1992); in this case a high concentration of glucose allows association of the different SNF1 subunits to form an active complex (Jian *et al.*, 1996). However, SNF1 is not only a glucose regulator in yeast, as it also has functions in sporulation,

glycogen accumulation and the biogenesis of peroxisomes (Hardy *et al.*, 1994; Simon *et al.*, 1992). Less is known about the functions of plant SNF1 related protein kinase (SnRK). SnRK activity in an extract of spinach leaf was able to phosphorylate and inactivate 3-HMG CoA reductase, nitrate reductase, and sucrose phosphate synthase *in vitro* (Sugden *et al.*, 1999). Antisense expression of *PKIN1* (SnRK1 from potato) results in the repression of sucrose synthase (Purcell *et al.*, 1998). While all functions of SnRKs are not fully understood, their role in the regulation of glucose metabolism, controlled by glucose 6- phosphate (Toroser *et al.*, 2000) is well-established. The function of *At*PTPKIS1 is then one of modulating this regulation of glucose metabolism (and probably of nitrate assimilation and lipid anabolism) by the SnRK complex, in response to changing cellular conditions. Evidence of this role was sought *in planta* by examining the phenotypes of plants containing knockouts in protein phosphatase genes, putatively including *At*PTPKIS1, but this did not give any positive result. More work to establish that the "knockout" had prevented *At*PTPKIS1 expression, and if necessary producing new mutants with true knockouts, would be necessary before any firm conclusion could be drawn on the effects of *At*PTPKIS1 on metabolic regulation.

Further work with *At*PTPKIS1 and its encoding gene would need to concentrate on the production of functional proteins, and on producing mutant lines where *At*PTPKIS1 expression would be assayed by analysis of mRNA. The KIS domain from *At*PTPKIS1, which can be expressed easily, could be used to raise antibodies, which could be used for protein analysis. The existence of protein complexes containing *At*PTPKIS1 in *A. thaliana* could then be demonstrated using immunoprecipitation (or a similar technique). The composition of complexes could be investigated *via* two-dimensional gel analysis, using non-denaturing conditions in the first dimension. Moreover, a confirmation *in planta* of the association of AKIN11 and *At*PTPKIS1 would be possible with suitable reporter gene constructs.

References

Allen, G.J. and Sanders, D. (1995) Calcineurin, a type 2B protein phosphatase, modulates the Ca^{2+} permeable slow vacuolar ion channel of stomatal guard cells. *Plant Cell* 7(9), pp. 1473-1483.

Barford, D. (1996) Molecular mechanisms of the protein serine threonine phosphatases. *Trends Biochem. Sci.* 21(11), pp. 407-412.

Beg, Z.H., Allmann, D.W., Gibson, D.M. (1973) Modulation of 3-hydroxy-3-methylglutaryl coenzyme A reductase activity with cAMP and with protein fractions of rat cytosol. *Biochem. Biophys. Res. Comm.* 54(4), pp. 1362-1369.

Bethk, P.C. and Jones, R.L. (1997) Reversible protein phosphorylation regulates the activity of the slow-vacuolar ion channel. *Plant J.* 11(6), pp. 1227-1235.

Bhalerao, R.P., Salchert, K., Bako, 1., Okresz, L., Szabados, L. Muranaka, T., Machida, Y., Schell, J., Koncz, C. (1999) Regulatory interaction of PRL1 WD protein with *Arabidopsis* SNF1-like protein kinase. *Proc. Natl. Acad. Sci. USA*, 96(9), pp. 5322-5327.

Bogre, L., Ligterink, W., Meskiene, 1., Barker, P.J., Heberle-Bors, E., Huskinsson, N.S., Hirt, H. (1997) Wounding induces the rapid and transient activation of a specific MAP kinase pathway. *Plant Cell* 9(1), pp. 75-83.

Camps, M., Nicholas, A., Arkinstall, S. (2000) Dual specificity phosphatases: a gene family for control of MAP kinase function. *FASEB J.* 14(1), pp. 6-16.

Carling, D., Zammit, V.A., Hardie, D.G. (1987) A common bi-cyclic protein kinase cascade inactivates the regulatory enzymes of fatty acid and cholesterol biosynthesis. *FEBS Lett.* 223(2), pp. 217-222.

Celenza, J.L., Carlson, M. (1986) A yeast gene that is essential for release from glucose repression encodes protein kinase. *Science* 233, pp. 1175-1180.

Chollet, R., Vidal, J., O'Leary, M.H. (1996). Phospho*enol*pyruvate carboxylase: a ubiquitous, highly regulated enzyme in plants. *Ann. Rev. Plant Phys.* 47, pp. 273-298.

Cori, G.T. and Cori, C.F. (1943) Crystalline muscle phophorylase IV. Formation of glycogen. *J. Biol. Chem.* 151(1), pp. 57-63.

Cori, G.T. and Green, A.A. (1943) Crystalline muscle phophorylase II. Prostetic group. *J. Biol. Chem.* 151(1), pp. 31-38.

Cohen, P. (1978) The role of cyclic-AMP-dependent protein kinase in the regulation of glycogen metabolism in mammalian skeletal muscle. *Curr. Top. Cell. Regul* 14, pp. 117-196.

Cohen, P. (1983) Protein phosphorylation and control of glycogen metabolism in skeletal muscle. *Philos. T. Roy. Soc. B* 302, pp. 13-25.

Cohen, P. (1989) The structure and regulation of protein phosphatases. *Ann. Rev. Biochem.* 58, pp. 453-508.

Cohen, P.T.W. (1997) Novel protein serine-threonine phosphatases: variety is the spice of life. *Trends Biochem. Sci.* 22(7), pp. 245-251.

Cregg, J.M., Vedvick, T.S., Raschke, W.C. (1993). Recent advances in the expression of foreign genes in *Pichia pastoris*. *Bio/Technology* 11, pp. 905-910.

Dale, S., Arró, M., Becerra, B., Morrice, N.G., Boronat, A., Hardie, DG, Ferrer, A (1995) Bacterial expression of the catalytic domain of 3-hydroxy-3-methylglutaryl-CoA re-ductase (isoform HMGR1) from *Arabidopsis thaliana*, and its inactivation by phosphorylation at Ser577 by *Brassicaoleracea* 3-hydroxy-3-methylglutaryl-CoA reductase kinase. *Eur. J. Biochem.* 233(2), pp. 506-513.

Das, A.K, Helps, N.R., Cohen, P.T.W., Bradford, D. (1996) Crystal structure of protein serine/threonine phosphatase 2C at 2.0 Å resolution. *EMBO J.* 15(24), pp. 6798-6809.

Davies, S.P., Carling, D., Hardie, D.G. (1989) Tissue distribution of the AMP-activated protein kinase, and lack of activation by cyclic-AMP-dependent protein kinase, studied using a specific and sensitive peptide assay. *Eur.J.Biochem.* 186(1-2), pp. 123-128.

Denu, J.M., Stuckey, J.A., Saper, M.A., Dixon, J.E. (1996) Form and function in protein dephosphorylation. *Cell* 87(3), pp. 361-364.

Denu, J.M.and Dixon, J.E. (1998) Protein tyrosine phosphatases: mechanisms of catalysis and regulation. *Curr. Opinion Chem. Biol.* 2(5), pp. 633-641.

Durfee, T., Becherer, K., Chen, P.L., Yeh, S.H., Yang, Y., Kilburn, A.E., Lee, W.H., Elledge, S.J. (1993) The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. *Genes Dev.* 7(4), pp. 555-569.

Egloff, M-P., Cohen, P.T.W., Reinemer, P., Barford, D. (1995) Crystal structure of the catalytic subunit of human protein phosphatase 1 and its complex with tungstate. *J. Mol. Biol.* 254(5), pp. 942-959.

Emanuelsson, O, Nielsen, H, von Heijne, G. (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* 8(5), pp. 978-984.

Emanuelsson O., Nielsen H, Brunak S., von Heijne G. (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence *J. Mol. B.* 300(4), pp. 1005-1016.

Ermolaeva, O., Rastogi, M., Pruitt, K.D., Schuler, G.D., Bittner, M.L., Chen, Y., Simon, R., Meltzer, P., Trent, J.M., Boguski, M.S. (1998) Data management and analysis for gene expression arrays. *Nat. Genet.* 20(1), pp. 19-23.

Farras, R, Ferrando, A, Jasik, J, Kleinow, T, Okresz, L, Tiburcio, A, Salchert, K, del Pozo, C, Schell, J, Koncz, C. (2001) SKP1-SnRK protein kinase interactions mediate proteasomal binding of a plant SCF ubiquitin ligase. *EMBO J.* 20(11), pp. 2742-2756.

Fauman, E.B. and Saper, M.A. (1996) Structure and function of the protein tyrosine phosphatases. *Trends Biochem Sci.* 21(11), pp. 413-417.

Faux, M.C., Scott, J.D. (1996) More on target with protein phosphorylation: conferring specificity by location. *Trends Biochem Sci.* 21(8), pp. 312-315.

Fernandez, A., Brautigan, D.L., Lamb, N.J. (1992) Protein phosphatase type 1 in mammalian cell mitosis: chromosomal localization and involvement in mitotic exit. *J. Cell Biol.* 116(6), pp. 1421-1430.

Ferrer, A, Caelles, C, Massot, N, Hegardt, F.G (1987) Affinity labelling of the catalytic and AMP allosteric sites of 3-hydroxy-3-methylglutaryl-coenzyme A reductase kinase by 5'-p-fluorosulfonylbenzoyladenosine. *J Biol Chem.* 262(28), pp. 13507-13512.

Field, S. and Song, O. (1989) A novel genetic system to detect protein protein interactions. *Nature* 340, pp. 245-246.

Fischer, E.H. and Krebs, E.G. (1955) Conversion of phosphorylase *b* to phosphorylase *a* in muscle extraction. *J. Biol. Chem.* 216(1), pp. 121-132.

Fordham-Skelton, A.P., Chilley, P., Lumbreras, V., Reignoux, S., Fenton, T.R., Dahm, C.C., Pages, M., Gatehouse, J.A. (2002) A novel higher plant protein tyrosine phosphatase interacts with SNF1-related kinases *via* a KIS (kinase interaction sequence) domain. *Pant J.* 29(6), pp. 705-715.

Fordham-Skelton, A.P., Skipsey, M., Evans, I.M., Edwards, R., Gatehouse, J.A. (1999) High plant tyrosine-specific protein phosphatases (PTPs) contain novel amino-terminal domains: expression during embryogenesis. *Plant Mol. Biol.* 39(3), pp. 593-605.

Gancedo, J.M. (1992) Carbon catabolite repression in yeast. *Eur. J. Biochem.* 206(2), pp. 297-313.

Garrington, T.P. and Johnson, G.L. (1999) Organization and regulation of mitogen-activated protein kinase signalling pathways. *Curr. Opinion Cell Biol.* 11(2), pp. 211-218.

Gebbink, M.F.B.G., Zondag, G.C.M., Wubbolts, R.W., Beijersbergen, R.L., Van Etten, I., Moolenaar, W.H. (1993) Cell-cell adhesion mediated by a receptor-like protein tyrosine phosphatase. *J. Biol. Chem.* 268(22), pp. 16101-16104.

Gillespie, J.G., Hardie, D.G. (1992) Phosphorylation and inactivation of HMG-CoA reductase at the AMP-activated protein kinase site in response to fructose treatment of rat hepatocytes. *FEBS Lett.* 306(1),59-62.

Green, A.A. and Cori, G.T. (1943) Crystalline muscle phophorylase I. Preparation, properties, and molecular weight. *J. Biol. Chem.* 151(1), pp. 21-29.

Gupta, R., Huang, Y., Kieber, J., Luan, S. (1998) Identification of a dual-specificity protein phosphatase that inactivates a MAP kinase from *Arabidopsis. Plant J.* 16(5), pp. 581-589.

Hardie, D.G. (1999) Plant protein serine/threonine kinases: classification and functions. *Annu. Rev. Plant Phys.* 50, pp. 97-131.

Hardie, D.G., Carling, D., Carlson, M. (1998) The AMP-activated /SNF1 protein kinase subfamily: metabolic sensors of the eucaryotic cell. *Annu. Rev. Biochem.* 67, pp. 821-855.

Hardy, T.A., Huang, D., Roach, P.J. (1994) Interaction between cAMP dependent and SNF1 protein kinases in the control of glycogen accumulation in *Saccharomyces cerevisiae. J. Biol. Chem.* 269(45), pp. 27907-27913.

Haring, M.A., Siderius, M., Jonak, C., Hirt, H., Watson, K.M., Musgrave, A. (1995) Tyrosine phosphatase signalling in a lower plant: cell-cycle and oxidative stress-regulated expression of the chlamydomonas eugaetos VH-PTP13 gene. *Plant J.* 7(6), pp. 981-988.

Harper, J.W., Adami, G.R., Wei, N., Keyomarsi, K., Elledge, S.J. (1993) The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell* 75(4), pp. 805-816.

Harper, J.F., Sussman, M.R., Schaller, G.E., Putnam-Evans, C., Charbonneau, H., Harmon, A.C. (1991) A calcium-dependent protein kinase with a regulatory domain similar to calmodulin. *Science* 252, pp. 951-954.

Hong, Y., Takano, M., Liu, C.M., Gasch, M.L., Chye, A., Chua, N.H. (1996) Expression of three members of the calcium-dependent protein kinase gene family in *Arabidopspis thaliana*. *Plant Mol. Biol.* 30(6), pp. 1259-1275.

Huang, J.M., Wei, Y.F., Kim, Y.H., Osterberg, L., Matthews H.R. (1991) Purification of a protein histidine kinase from the yeast Saccharomyces cerevisiae. The first member of this class of protein kinases. *J. Biol. Chem.* 266(14), pp. 9023-9023.

Huang, Y, Li H., Gupta, R., Morris, P., Luan, S., Kieber, J.J. (2000) *At*MPK4 an *Arabidopsis* homologue of mitogen-activated protein kinase, is activated *in vitro* by *At*MEK1 through threonine phosphorylation. *Plant Physio.* 122(4), pp. 1301-1310.

Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* 3(3), pp. 318-356.

Jencks, W.P. (1989) How does a calcium pump pump calcium? *J. Biol. Chem.* 264(32), pp. 18855-18858.

Jiang, R and Carlson, M. (1996) Glucose regulates protein interactions within the yeast SNF1 protein kinase complex. *Genes Dev.* 10(24), pp. 3105-3115.

Jiang, R. and Carlson, M. (1997) The Snf1 protein kinase and its activating subunit, Snf4, interact with distinct domains of the Sip1/Sip2/Gal83 component in the kinase complex. *Mol. Cell. Biol.* 17(4), pp. 2099-2106.

Jiang, Y.P., Wang, H., D'Eustachio, P., Musacchio, J.M, Schlesinger, J., Sap, J. (1993)Clonning and characterisation of R-PTP-KAPPA, a new member of the receptor protein tyrosine phosphatase family with a proteolytically cleaved cellular adhesion molecule-like extracellular region. *Mol. Cell. Boil.* 13(5), 2942-2951.

Johnson, L.N., Snape, P., Martin, J.L., Acharya, K.R., Barford, D., Oikonomako, N.G. (1993) Crytallographic binding studies on the allosteric inhbitor glucode-6-phosphtate to T state glycogen phosphorylase B. *J. Mol. Biol.* 232(1), pp. 253-267.

Jonack, C., Ökrész, L., Bögre, L., Hirt, H. (2002) Complexity cross talk and integration of plant MAP kinase signalling. *Current Opinion in Plant Biology*, 5(5), pp. 415-424.

Kaiser, W.M. and Huber, S.C. (2001) Post-translational regulation of nitrate reductase: mechanism, physiological relevance and environmental triggers. *J. Exp. Bot.* 52(363), pp. 1981-1989.

Keith, M.L, Rodwell, V.W, Rogers, D.H, Rudney, H. (1979) *In vitro* phosphorylation of 3-hydroxy-3-methylglutaryl coenzyme A reductase: analysis of 32P-labeled, inactivated enzyme. *Biochem. Bioph. Res. Co.* 90(3), pp. 969-975.

Klee, C.B., Draetta, G.F., Hubbard, M.J. (1988) Calcineurin. In: Meister, A. *Adv. Enzymology* 61 pp. 149-200.

Knetsch, M.L.W., Wang, M., Snaarjagalska, B.E., Heimovaaradijkstra, S. (1996) Acid-acid induces mitogen-activated protein-kinase activation in barley aleurone protoplasts. *Plant Cell* 8(6), pp. 1061-1067.

Krebs, E.G. and Beavo, J.A. (1979) Phosphorylation-Dephosphorylation of Enzymes *Ann. Rev. Biochem.* 48, pp. 923-959.

Kreis, M. and Walker, J.C. (2000) *Plant protein kinases.* London: Academic Press.

Kroj, T., Rudd, J.J., Nürnberger, T., Gäbler, Y., Lee, J., Scheel, D. (2003) Mitogene-activated protein kinases play an essential role in oxidative burst-

independent expression of pathogenesis-related genes in parsley. *J. Biol. Chem.* 278(4), pp. 2256-2264.

Kudla, J, Xu, Q, Harter, K, Gruissem, W, Luan, S. (1999) Genes for calcineurin B-like proteins in *Arabidopsis* are differentially regulated by stress signals. *P. Natl. Acad. Sci. USA* 96(8), pp. 4718-4723.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227(259), pp. 680-685.

Lee, J.Y., Yoo, B.C., Harmon, A.C. (1998) Kinetic and calcium-binding properties of three calcium-dependent protein kinase isoenzymes from soybean. *Biochemistry* 37(19), pp. 6801-6809.

Levine, R., and Ginsburgh, A. (1985) Modulation by molecular interactions. *Curr. Top. Cell. Regul.* 26, pp. 177-346.

Li, C. and Xu, Q. (2000) Mechanical stress-initiated signal transductions in vascular smooth muscle cells. *Cell. Signal.* 12(7), pp. 435-445.

Li, W, Luan, S, Schreiber, S.L, Assmann, S.M. (1994) Evidence for protein phosphatase 1 and 2A regulation of K+ channels in two types of leaf cells. *Plant physiol.* 106(3), pp. 963-970.

Li, Y.M, Casida, J.E. (1992) Cantharidin-binding protein: identification as protein phosphatase 2A. *P. Natl. Acad. Sci. USA* 89(24), pp. 11867-11870.

Liu, J, Zhu, J.K. (1998) A calcium sensor homologue required for plant salt tolerance. *Science* 280(5371), pp. 1943-1945.

Liu, Q., Zhang, Y., Chen, S. (2000) Plant protein kinase genes induced by drought, high salt and cold stresses. *Chinese Sci. Bull.* 45(13), pp. 1153-1157.

Loh, C, Carew, J.A, Kim, J, Hogan, P.G, Rao, A. (1996) T-cell receptor stimulation elicits an early phase of activation and a later phase of deactivation of the transcription factor NFAT1. *Mol. Cell. Biol.* 16(7), pp. 3945-3954.

López-Bucio, J., Hernández-Abreu, E., Sánchez-Calderón, L., Nieto-Jacobo, M.F., Simpson, J., Herrera-Estrella, L. (2002) Phosphate availability alters architecture and causes changes in hormone sensitivity in *Arabidopsis* root system. *Plant physiol.* 129(1), pp. 244-256.

Luan, S, Li, W, Rusnak, F, Assmann, S.M, Schreiber, S.L. (1993) Immunosupressants implicate protein phosphatase regulation of K^+ channel in guard cells. *Proc. Natl. Acad. Sci. USA* 90(6), pp. 2202-2206.

Luan, S. (1998) Protein phosphatases and signaling cascades in higher plants. *Trends Plant Sci.* 3(7), pp. 271-275.

Luan, S. (2002) Signaling drought in guard cells. *Plant Cell Envir.* 25(2), pp. 229-237.

Ludin, K., Jiang, R., Carlson, M. (1998) Glucose-regulated interaction of a regulatory subunit of protein phosphatase 1 with the Snf1 protein kinase in *Sacchatomyces cerevisiae. Proc. Natl. Acad. Sci.USA* 95(11), pp. 6245-6250.

Lumbreras, V., Alba, M.M., Kleinow, T., Koncs, C., Pages, M. (2001) Domain fusion between SNF1-related kinase subunits during plant evolution. *EMBO Report* 2(1), pp. 55-60.

MacKintosh, C. (1992) Regulation of spinach-leaf nitrate reductase by reversible phosphorylation. *Biochem. Biophys Acta* 1137(1), pp. 121-126.

MacKintosh, C., Lyon, G.D., MacKintosh, R.W. (1994) Protein phosphatase inhibitors activate anti-fungal defence responses of soybean cotyledons and cell cultures. *Plant J.* 5(1), pp. 137-147.

Martin, T., Oswald, O., Graham, I.A. (2002) Arabidopsis seedling growth, storage lipid mobilisation, and photosynthetic gene expression are regulated by carbon:nitrogen availability. *Plant Physiol.* 128(2), pp. 472-481.

Meskienne, I., Bögre, L, Glaser, W., Balog, J., Brandstötter, M., Zwerger, K., Ammerer, G., Hirt, H. (1998) MP2C, a plant protein phosphatase 2C, functions as a negative regulator of mitogen-activated protein kinase pathways in yeast and plants. *Proc. Natl. Acad. Sci. USA* 95(4), pp. 1938-1943.

Minassian, B.A., Lee, J.R., Herbrick, J.A., Huizenga, J., Soder S., Mungall, A.J., Dunham, I., Gardner, R., Fong, C.Y., Carpenter, S., Jardim, L., Satischandra, P., Andermann, E., Snead, O.C., Lopes-Cendres, I., Tsui, L.C., Delgado-Escueta, A.V., Rouleau, G.A., Scherer, S.W. (1998) Mutation in a gene encoding a novel protein tyrosine phosphatase cause progressive myoclonus epilepsy. *Nature Genet.* 20(2), pp. 171-174.

Mizoguchi, T., Gotoh, Y., Nishida, E., Yamaguchi-Shinozaki, K., Hayashida, N., Iwasaki, T., Kamada, H., Shinozaki, K. (1994) Characterization of two cDNA that encode MAP kinase homologues in *Arabidopsis thaliana* and analysis of the possible role of auxin in activating such kinase activities in cultured cells. *Plant J.* 5(1), pp. 111-122.

Monod, J., Changeux, J.-P., Jacob, F. (1963) Allosteric proteins and molecular systems. *J. Mol. Biol.* 6(4), pp. 306-329.

Nakai, K. and Kanehisa, M. (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics* 14(4), pp. 897-911.

Nemeth, K., Salchert, K., Putnoky, P., Bhalerao, R., Koncz-Kalman, Z., Stankovic-Stangeland, B., Bako, L., Mathur, J., Okresz, L., Stabel, S., Geigenberger, P., Stitt, M., Redei, G.P., Schell, J., Koncz, C. (1998) Pleiotropic control of glucose and hormone responses by PRL1, a nuclear WD protein, in *Arabidopsis*. *Genes Dev.* 12(19), pp. 3059-3073.

Neurath, H. (1986) The versatility of proteolytic enzymes. *J. Cell. Biochem.* 32(1), pp. 35-49.

Odjavkova, M. and Hadjiivanova C. (2001) The complexity of pathogen defense in plants. *Bulg. J. Plant Physiol.* 27(1-2) pp. 101-109.

O'Malley, B.W., Towle, H.C., Schwartz, R.J. (1977) Regulation of Gene Expression in Eucaryotes. *Annu. Rev. Genet.* 11, pp. 239-275.

Pardee, A.B. (1959) The control of enzyme activity. In: Boyer, P.D., Lardy, H., Myrbäck, K. *The Enzymes.* Second edition, London: Academic Press, pp. 681-716.

Pawson, T. (1995) Protein modules and signalling networks. *Nature* 373, pp. 573-580.

Pelham, H.R.B., Jackson, R.J. (1976) An efficient mRNA-dependent translation system from retyculocyte lysates. *Eur. J. Biochem.* 67(1), pp. 247-256.

Purcell, P.C., Smith, A.M., Halford, N.D. (1998) Antisense expression of a sucrose nonfermentating-1-related protein kinase sequence in potato results in decreased expression of sucrose synthase transcripts in leaves. *Plant J.* 14(2), pp. 195-202.Raz, V, Fluhr, R. (1993) Ethylene Signal Is Transduced *via* Protein Phosphorylation Events in Plants. *Plant Cell* 5(5), pp. 523-530.

Sambrook, J., Russell, D.W. (2001) *Molecular Cloning – a laboratory manua.l* 3rd edition Cold Spring Harbour Laboratory.

Schaffer, R., Landgraf, J., Accerbi, M., Simon, V., Larson, M., Wisman, E. (2001) Microarray analysis of diurnal and circadian-regulated genes in *Arabidopsis. Plant Cell* 13(1), pp. 113-123.

Schena, M., Shalon, D., Davis, R.W., Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA Microarray. *Science* 270(5235), pp. 467-470.

Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O., Davis, R.W. (1996) Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci. USA* 93(20), pp. 10614-10619.

Schiiller, H.J., and Entian, K.D. (1987) Molecular characterisation of yeast regulatory gene Cat3 necessary for glucose depression and nuclear localisation of its product. *Gene* 67(2), pp. 247-257.

Shi, J., Kim, K-N., Ritz, O., Albrecht, V., Gupta, R., Harter, K., Luan, S., Kudla, J. (1999) Novel Protein Kinases Associated with Calcineurin B-like Calcium Sensors in *Arabidopsis. Plant Cell* 11(12), pp. 2393-2405.

Shifrin, V.I., Davis, R.J., Neel, B.G. (1997) Phosphorylation of Protein-tyrosine phosphatase PTP-1B on identical sites activation of a common signalling pathway during mitosis and stress response in mammalian cells. *J. Biol. Chem.* 272 (5), pp. 2957-2962.

Simon, M., Binder, M., Adam, G., Hartig, A., Ruis, H. (1992) Control of peroxisome proliferation in *Saccharomyces cerevisiae* by *ADR1*, *SNF1* (*CAT1, CCR1*) and *SNF4* (*CAT3*). *Yeast* 8(4), pp. 303-309.

Smith, D.B., Johnson, K.S. (1988) Single-step purification of polypeptides in *Escherichia coli* as fusions with glutathione S-transferase. *Gene* 67(1), pp. 31-40.

Smith, R.D., Walker, J.C. (1996) Plant protein phosphatases. *Ann. Rev. Plant Phys. Plant Mol. Biol.* 47, pp. 101-125.

Sprang, S.R., Acharya, K.R., Goldsmith, E.J., Stuart, D.I., Varvill, K., Fletterick, R.J., Madsen, N.B., Johnson, L.N. (1988) Sturcyural changes in glycogen phosphorylase induced by phosphorylation. *Nature* 336, pp. 215-221.

Stadtman, E.R. and Chock, P.B. (1978) Interconvertible enzyme cascades in metabolic regulation. *Curr. Top. Cell. Regul.* 13, pp. 53-95.

Stone, J.M., and Walker, J.C. (1995) Plant protein kinase families and signal transduction. *Plant Physiol.* 108(2), pp.451-457.

Sugden, C., Donaghy, P.G., Halford, N.G., Hardie, D.G. (1999) Two SNF1-related protein kinases from spinach leaf phosphorylate and inactivate 3-hydroxy-3-methylglutaryl-coenzyme A reductase, nitrate reductase, and sucrose phosphate synthase *in vitro*. *Plant Physiol* 120(1), pp. 257-274.

Suzuki, K., Shinshi, H. (1995) Trascient activation and tyrosine phosphorylation of a protein kinase in tabaco cells treated with a fungal elicitor. *Plant Cell* 7(5), pp. 639-647.

Thompson, J., Francois, J., Gaugran, J.P., Tatchel, K (1991) Deletion of SNF1 affects the nutrient response of yeast and resembles mutations which activates the adenylate cyclase pathway. *Genetics* 129(3), pp. 697-706.

Toroser, D., Paut, Z., Hubert, S.C. (2000) Regulation of plant SNF1-Related protein kinase by glucose-6-phosphate. *Plant Physiol.* 123(1), pp. 403-411.

Ulm, R, Ichimura, K., Mizoguchi, T., Peck, C.S., Zhu, T., Wang, X., Shinozaki, K., Paszkowski, J. (2002) Dinstinct regulation of salinity and genotoxic stress responses by *Arabidopsis* MAP kinase phosphatase 1. *EMBO J.* 21(23), pp. 6483-6493.

Ulm, R., Revenkova, E., di Sansebastiano, G.P., Bechtold, N., Paszkowski, J. (2001) Mitogene-activated protein kinase phosophatase is required for genotoxic stress relief in *Arabidopsis*. *Genes Dev.* 15(6), pp. 699-709.

Waterham, H.R., Digan, M.E., Koutz, P.J., Lair, S.V., Cregg, J.M. (1997) Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. *Gene* 186(1) pp. 37-44.

Wera, S, Hemmings, BA. (1995) Serine/threonine protein phosphatases. *Biochem J.* 311(1), pp. 17-29.

Winter, H., Huber, S.C. (2000) Regulation of sucrose metabolism in higher plants: localization and regulation of activity of key enzymes. *Crit. Rev. Biochem. Mol.* 35(4), pp. 253-289.

Woods, A, Munday, M.R., Scott, J., Yang, Y.L., Carlson, M., Carling, D. (1994) Yeast SNF1 is functionally related to mammalian AMP-activated protein kinase and regulates acetyl-CoA carboxylase *in vivo*. *J. boil. Chem.* 269(30) pp. 19509-19515.

Woods, A., Salt, I., Scott, J., Hardie, D.G., Carling, D (1996) The alpha1 and alpha2 isoforms of the AMP-activated protein kinase have similar activities in rat liver but exhibit differences in substrate specificity in vitro. *FEBS Lett.* 397(2-3), pp. 347-351.

Xu, Q., Fu, H.H., Gupta, R., Luan, S. (1998) Molecular cloning and characterisation of a tyrosine-specific protein phosphatase encoded by a stress-responsive in *Arabidopsis*. *Plant Cell* 10(5), pp. 849-857.

Yamamoto T, Takeda K, Harada S, Nakata T, Azuma A, Sasaki S, Akagawa M. (2003) HMG-CoA reductase inhibitor enhances inducible nitric oxide synthase expression in rat vascular smooth muscle cells; involvement of the Rho/Rho kinase pathway. *Atherosclerosis* 166(2), pp. 213-222.

Yang, K.-Y., Liu, Y., Zhang, S. (2001) Activation of a mitogene-activated protein kinase pathway is involved in disease resistance in tobacco. *Proc. Natl. Acad. Scio USA* 98(2), pp. 741-746.

Yanovsky, M. (2001) Signaling networks in the plant circadian system. *Curr. Opini. Plant Biol.* 4(5), pp. 429-435.

Yu, S.M. (1999) Cellular and genetic responses of plants to sugar starvation. *Plant Physiol.* 121(3), pp. 687-693.

Zhang, H., Forde, B.G. (1998) An *Arabidopsis* MADS box gene that controls nutrient-induced changes in root architecture. *Science* 279(5349), pp. 407-409.

Zhang, S.Q., Klessig, D.F. (1997) Salicylic-acid activates a 48-KD MAP kinase in tabacco. *Plant Cell* 9(5), pp. 809-824.