

Ligands of Inositol Polyphosphate Binding Proteins

A thesis submitted to the University of East Anglia for the degree of
Doctor of Philosophy

by

Megan Gilmartin

December 2022

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived therefrom must be in accordance with current UK Copyright Law.

In addition, any quotation or extract must include full attribution.

Abstract

In plants, cytosolic and organellar inositol polyphosphate (InsP) molecules are involved in multiple stress response pathways, such as drought tolerance and herbivory defence. The turnover and function of InsPs *in vivo* depend on enzymes including inositol phosphate kinases (IPKs) and phosphatases. InsP receptor proteins co-ordinate cellular responses dependent on the pool of InsPs present. Many IPKs localize to the nucleus, including AtIPK1 which is responsible for the synthesis of phytic acid, the implicated regulatory ligand of an mRNA transporter complex, a phytohormone co-receptor complex and precursor molecule to a subclass of InsPs, inositol pyrophosphates (PP-InsPs). Despite implications of IPKs in nuclear regulatory roles, these have not been described in plants. The primary research aim was to investigate a nuclear role of AtIPK1 through DNA binding assays (Chapter 2) and structure-function studies (Chapter 3).

SPX domains have recently been identified as PP-InsP ligand receptors which mediate plant phosphate homeostasis. A secondary aim of this project was to characterize a subset of Arabidopsis SPX domain proteins (AtSPX1, AtSPX3, AtPHT5;1), with a particular focus on substrate discrimination towards InsPs and PP-InsPs. To this end protein production, ligand-binding assays and structure-function studies were attempted (Chapter 4).

This thesis presents evidence for novel DNA binding of AtIPK1 and AtSPX1, whereby DNA binding activity is regulated by InsP titrations. In the application of *in silico* analyses, potential DNA binding sites were identified whilst low structural homology between these proteins and known DNA binding proteins were observed. A novel structure of wildtype apo AtIPK1 was solved lending new insights into conformational changes upon ligand binding. Additionally, AtSPX1 ligand binding was studied showing subtle differences of InsP preference in the absence of DNA interaction partner. Thus, demonstrating how InsP-binding proteins may exhibit additional regulatory functions which depend upon specific InsP substrate interactions.

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate 'take down' action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

Table of contents

Abstract	2
Table of contents	3
List of Tables	8
List of Figures	9
Acknowledgements	12
1 Introduction	13
1.1 <i>Nomenclature of inositol phosphates</i>	13
1.2 <i>Biosynthetic pathways of InsPs and PP-InsPs in plants</i>	14
1.3 <i>Pi, InsP and PP-InsP cycling</i>	17
1.4 <i>Roles in abiotic stress response</i>	19
1.5 <i>Modulation of signalling pathways through InsP binding</i>	20
1.6 <i>Regulation of nuclear processes</i>	22
1.7 <i>Aims of project</i>	23
2 AtIPK1 interacts with DNA <i>in vitro</i>	25
2.1 <i>Introduction</i>	25
2.1.1 General characterization of AtIPK1	25
2.1.2 Investigative tools for DNA-protein interactions	27
2.1.3 Aims of experiments	37
2.2 <i>Material and methods</i>	38
2.2.1 Expression and purification of AtIPK1	38
2.2.1.1 Propagation of <i>pOPINF[T7lacO/AtIPK1]</i> in Stellar	38
2.2.1.2 DNA gel electrophoresis	38
2.2.1.3 Heat shock transformations into competent cells	38
2.2.1.4 Glycerol stocks for <i>pOPINF [T7lacO/AtIPK1]</i>	39
2.2.1.5 Large scale <i>AtIPK1</i> expression and cell harvest	39
2.2.1.6 French pressure cell	40

2.2.1.7	Purification of AtIPK1	40
2.2.1.8	SDS-PAGE analysis of cell pellets, supernatant and purified AtIPK1	42
2.2.2	Protein-DNA binding assays	43
2.2.2.1	Oligomer design and synthesis.....	43
2.2.2.2	Fluorescence polarization with AtIPK1	44
2.2.2.3	F-EMSA.....	45
2.2.2.4	Analytical ultracentrifuge sedimentation equilibrium	46
2.3	<i>Results</i>	51
2.3.2	AtIPK1 binds to DNA in FP assays	52
2.3.3	F-EMSA	56
2.3.4	AtIPK1 is a dimer in solution	59
2.3.5	Preliminary data from FLIM-FRET	64
2.4	<i>Discussion</i>	68
2.4.1	AtIPK1 binds DNA <i>in vitro</i>	68
2.4.2	Non-specific DNA binding of AtIPK1.....	70
2.4.3	Evidence of IPK dimer formation.....	72
2.4.4	Towards <i>in vivo</i> evidence	74
2.5	<i>Conclusion</i>	76
3	AtIPK1 structure and mutagenesis	78
3.1	<i>Introduction</i>	78
3.1.1	Structural characterization of AtIPK1	78
3.1.2	AtIPK1 substrate specificity and conformation changes	79
3.1.3	Putative DNA binding motifs of AtIPK1	82
3.1.4	Aims of experiments	84
3.2	<i>Materials and methods</i>	86
3.2.1	Crystallography trials with AtIPK1 and dsDNA	86
3.2.1.1	Crystallography screens with AtIPK1.....	86
3.2.1.2	Crystal optimization, harvest and data collection	86
3.2.1.3	Data processing and refinement.....	87
3.2.2	SDM of AtIPK1	87
3.2.2.1	Primer design and PCR.....	87
3.2.2.2	In-fusion cloning AtIPK1 mutants into pOPINF	89
3.2.2.3	Transformation into competent <i>E. coli</i> strains.....	90

3.2.3.1	Expression trials of AtIPK1 GRP mutants.....	90
3.2.3.2	Solubility test of AtIPK1 SDM mutants	91
3.2.3.3	Large scale protein production of AtIPK1 GRP mutants.....	91
3.2.3.4	Purification of AtIPK1 GRP mutants.....	91
3.2.4	SDS-PAGE analysis.....	91
3.2.5	HPLC assays.....	92
3.2.6	FP assays.....	92
3.2.6.1	Wildtype AtIPK1 vs SDM mutants.....	92
3.2.6.2	AtIPK1-DNA displacement assays	93
3.2.7	F-EMSA with wildtype AtIPK1 vs SDM mutants.....	93
3.2.8	<i>In silico</i> investigation of AtIPK1.....	93
3.3	<i>Results</i>	95
3.3.1	AtIPK1 co-purifies with nucleotide	95
3.3.2	Novel apo AtIPK1 structure	98
3.3.3	Comparison of AtIPK1 structure models	99
3.3.4	Investigating the AtIPK1 GRP motif.....	103
3.3.4.1	Production of active AtIPK1 GRP mutants.....	103
3.3.4.2	AtIPK1 GRP mutant kinase activity	106
3.3.5	AtIPK1 GRP mutants binding DNA.....	110
3.3.6	Conformational changes of AtIPK1 and ligand binding.....	113
3.3.7	Investigating AtIPK1-DNA binding <i>in silico</i>	114
3.4	<i>Discussion</i>	120
3.4.1	Crystallization of protein-DNA complex.....	120
3.4.2	ADP and apo models of wildtype AtIPK1	122
3.4.3	Investigation of the AtIPK1 GRP motif	124
3.4.4	Regulation of AtIPK1-DNA binding activity.....	126
3.4.5	<i>In silico</i> DNA docking of AtIPK1	128
3.5	<i>Conclusion</i>	130
4	SPX domain proteins in Arabidopsis.....	132
4.1	<i>Introduction</i>	132
4.1.1	SPX domains are Pi sensors.....	132
4.1.2	Characterization of plant SPX domain containing proteins	134
4.1.3	AtSPX1 is an InsP ₈ receptor	136

4.1.4	Experimental aims.....	140
4.2	<i>Materials and methods</i>	142
4.2.1	Cloning and expression of SPX genes in <i>E. coli</i>	142
4.2.1.1	In-fusion cloning SPX genes into pOPINF	142
4.2.1.2	Expression trials and solubility testing of SPX gene products	142
4.2.1.3	Large scale production of full length AtSPX1 and AtSPX3	143
4.2.1.4	Purification of SPX proteins.....	143
4.2.2	Cloning SPX domains into <i>Pichia pastoris</i>	144
4.2.2.1	Cloning SPX domains into pGAPZ α A	144
4.2.2.2	Preparing competent <i>Pichia pastoris</i>	147
4.2.2.3	Transformation of SPX constructs into <i>Pichia pastoris</i>	148
4.2.2.4	Small scale expression trials in <i>Pichia</i>	148
4.2.3	Characterization of SPX proteins.....	148
4.2.3.1	FP with AtSPX1.....	148
4.2.3.2	Crystallography screens with AtSPX1.....	149
4.2.3.3	F-EMSA with AtSPX1	150
4.2.3.4	Modelling SPX domain proteins.....	150
4.3	<i>Results</i>	151
4.3.1	Cloning, expression and purification of SPX proteins.....	151
4.3.1.1	Cloning and expression of SPX domains in <i>E. coli</i>	151
4.3.1.2	Cloning and expression of SPX domains in <i>Pichia pastoris</i>	152
4.3.1.3	Large scale production of AtSPX1 in <i>E. coli</i>	153
4.3.2	Binding of InsPs and PP-InsPs by AtSPX1	155
4.3.2.1	Investigation of AtSPX1 ligand binding with FP.....	155
4.3.2.2	Investigation of AtSPX1 binding preference of InsPs and PP-InsPs..	156
4.3.3	Interaction of AtSPX1 with DNA.....	158
4.3.3.1	AtSPX1 binds different FAM-labelled DNA probes.....	158
4.3.3.2	AtSPX1-DNA-P ligand interactions	160
4.3.4	Structure predictions of SPX proteins	162
4.4	<i>Discussion</i>	165
4.4.1	Production of Arabidopsis SPX containing protein.....	165
4.4.1.1	Expression of SPX domain containing proteins	166
4.4.1.2	Purification of SPX proteins.....	167
4.4.2	AtSPX1 InsP and PP-InsP ligand binding	168

4.4.3	AtSPX1 interaction with DNA and InsPs <i>in vitro</i>	170
4.4.4	Implications of AtSPX1 DNA binding activity.....	172
4.5	<i>Conclusion</i>	173
5	General conclusions and future work	174
	Appendices	188
	Abbreviations	230
	References	234

List of Tables

Table 2-1 DNA oligomers synthesized by Eurofins.....	44
Table 2-2 Buffers used for optimization trial in SE experiments with 10 μ M AtIPK1	47
Table 2-3 The calculated average molecular mass of AtIPK1 from SE	59
Table 2-4 SE analysis of 10 μ M AtIPK1 in different buffering solutions	61
Table 2-5 Calculated molecular mass determined from SE with AtIPK1.....	62
Table 2-6 Calculated molecular mass determined from SE analysis of AtIPK1 and DNA in different micromolecular ratios	63
Table 2-7 Calculated molecular mass determined from SE analysis of AtIPK1 and DNA in different micromolecular ratios	64
Table 3-1 Primers for SDM of AtIPK1	88
Table 3-2 AtIPK1 primers for In-fusion into pOPINF	89
Table 3-3 Crystal data collection, structure determination, and refinement	96
Table 3-4 Summary of expression data for pOPINF[T7lacO:AtIPK1] GRP mutants	105
Table 4-1 Primers designed for In-fusion cloning full length SPX domains	142
Table 4-2 Primers designed for two methods of cloning SPX domains into pGAPZ α A.....	146
Table 4-3 Sequencing primers for validation of SPX domain cloning.....	147
Table 4-4 5'-FAM-labelled double stranded oligomers used in FP with AtSPX1.	149
Table 4-5 Inositol pyrophosphate ligands used in FP	149
Table 4-6 Summary of expression data of pOPINF[T7lacO]-SPX constructs in different <i>E. coli</i> strains	152

List of Figures

Figure 1.1 The structure of InsP ₆ and Agranoff's turtle	14
Figure 1.2 Biosynthesis of InsPs and PP-InsPs in plants.....	15
Figure 2.1 Diagrammatic representation of fluorescence polarization of unbound fluorescent probe and probe in complex	29
Figure 2.2 Diagrammatic representation of F-EMSA gel set-up	30
Figure 2.3 Representation of absorbance scan data from protein in an AUC experiment	32
Figure 2.4 Simple illustration of FRET principles	34
Figure 2.5 Fluorescent lifetime imaging microscopy.....	36
Figure 2.6 Example of SDS-PAGE analysis of protein samples from AtIPK1 purification.....	52
Figure 2.7 AtIPK1 binds to FAM-labelled DNA in FP assay	53
Figure 2.8 Salt dependence of AtIPK1-DNA interaction monitored by FP with FAM-AT 40mer.....	54
Figure 2.9 Annealed nucleotide oligomers run on 20% polyacrylamide gel.	55
Figure 2.10 AtIPK1 binds to three different FAM-oligomers.....	56
Figure 2.11 F-EMSA gel analysis of AtIPK1 interaction with DNA.....	58
Figure 2.12 SDS-PAGE analysis of AtIPK1 stability	60
Figure 2.13 Two representative SE curves for AtIPK1 in solution	61
Figure 2.14 Microscopy images of GFP-transformed <i>N. benthamiana</i> leaf discs stained with SO.....	66
Figure 2.15 Distributions of average lifetime (τ) measurements of GPF and AtIPK1-GFP transiently expressed in <i>N. benthamiana</i> leaf nuclei.....	67
Figure 3.1 A 2.02 Å cartoon representation of AtIPK1 in complex with InsP ₆ and ADP	79
Figure 3.2 Superimposed models of the three conformations of AtIPK1.....	81
Figure 3.3 An expanded view of the AtIPK1 putative DNA binding motif	83
Figure 3.4 Cartoon representation of the 2.23 Å AtIPK1-ADP structure	97
Figure 3.5 Two orientations of the superposition of wildtype apo AtIPK1 model against ADP bound AtIPK1 model	99
Figure 3.6 Superposition of native apo AtIPK1 model against the apo W129A model and ADP-bound native AtIPK1.....	101
Figure 3.7 Alignments of alike wildtype and W129A mutant structures	102

Figure 3.8 Swiss models of wildtype AtIPK1 and SDM mutants.....	103
Figure 3.9 Verification of SDM AtIPK1 by gel electrophoresis.....	104
Figure 3.10 SDS-PAGE results of representative expression trial for AtIPK1 SDM mutants in <i>E. coli</i> SHuffle T7 LysY	105
Figure 3.11 Example SDS-PAGE analysis of AtIPK1 GKP purification.....	106
Figure 3.12 SDS-PAGE analysis of purified AtIPK1 SDM mutants alongside wt on a 12% polyacrylamide gel.....	107
Figure 3.13 Activity of AtIPK1 SDM mutants monitored by ATP to ADP conversion using HPLC	108
Figure 3.14 Wildtype AtIPK1 and SDM mutants bind to 2-FAM-InsP ₅ monitored by FP assay	110
Figure 3.15 Wildtype AtIPK1 and SDM mutants bind to FAM-12mer	112
Figure 3.16 F-EMSA analysis with wildtype vs SDM mutants of AtIPK1 using FAM-12mer on an 8% NATIVE gel	113
Figure 3.17 Displacement of the AtIPK1-DNA complex by different ligands using FP with 500 nM protein and 2 nM DNA.....	114
Figure 3.18 Predicted positive binding surface of (a) AtIPK1 and (b) AtITPK4 from BINDup, with (c) additional features annotated for AtIPK1.....	116
Figure 3.19 Alignment of GRP hook peptide (in complex with duplex 12mer) and WT AtIPK1-ADP.	117
Figure 3.20 Two disparate models of DNA docking to wildtype apo AtIPK1 generated from the HDOCK webserver.....	118
Figure 3.21 Model of AtIPK1 homodimer from GalaxyHomomer	119
Figure 4.1 The SPX domain structural motif with conserved SPX fingerprint residues.....	133
Figure 4.2 Schematic examples of plant SPX domain proteins with different C-terminal domains modulating Pi homeostasis	135
Figure 4.3 Model for AtSPX1 regulation of AtPHR1 PSR promoted by 1,5-InsP ₈ . ..	139
Figure 4.4 SDS-PAGE analysis of AtSPX1 on 12% polyacrylamide gels	154
Figure 4.5 FP with high and low A260/280 ratio preps of AtSPX1 binding.....	156
Figure 4.6 Inositol polyphosphates and inositol pyrophosphates used in this study	157
Figure 4.7 Displacement of AtSPX1 binding of 2-FAM-InsP ₅ with InsPs and PP-InsPs.....	158
Figure 4.8 AtSPX1 binding different 5'FAM-labelled oligomers	159

Figure 4.9 Displacement of AtSPX1 binding 2 nM FAM-GC 40mer with InsPs and PP-InsPs.....	161
Figure 4.10 Phyre2 model of AtSPX1 aligned with SPX^{CtVtc4} in complex with InsP₆ in PyMOL.....	163
Figure 4.11 Phyre2 model of AtSPX1 with putative DNA binding residues and InsP₆ placement.....	164
Figure 5.1 Model for AtSPX1 directly and indirectly regulating AtPHR1	178
Figure 5.2 A simplified schematic of InsP metabolism and transport	181

Acknowledgements

My sincere thanks:

To my colleagues and mentors at UEA, who patiently answered my questions and helped me along.

To my friends and family, who continually supported and loved me through the many times I couldn't find the words.

To Hayley, who fits in categories above but also persevered with me through writing and preparing this thesis to the very end.

To Charles, who supported any of my choices and helped review this thesis.

To J, who was always with me, through happy times and the bad too.

1 Introduction

The inositol polyphosphates (InsPs) are a numerous and diverse group of water-soluble molecules synthesized in eukaryotic cells. They have been studied traditionally within a context of Pi storage, as phytate in plants; calcium-mediated signalling cascades in animals; immune responses of animals and plants; nuclear processing regulation in yeast, plants and animals; and Pi sensing in yeast, plants and animals. In plants, InsPs are thought to co-ordinate a variety of physiological responses to biotic and abiotic stresses. Within this thesis, particular attention is given to the characterization of proteins that bind these phosphate rich substrates/ligands within the plant model organism, *Arabidopsis thaliana*.

1.1 Nomenclature of inositol phosphates

Inositol is a cyclohexane ring with hydroxyl groups on each carbon and has nine stereoisomeric forms (Thomas *et al.*, 2016). The most abundant of which, *myo*-inositol (also denoted Ins or I), has an axially-oriented 2-hydroxyl, relative to the 'plane' of the puckered carbon ring (Parthasarathy and Eisenberg, 1986), with all other hydroxyls in equatorial positions (Figure 1.1a). The inositol phosphates (InsPs or IPs) are inositol molecules with one or more single phosphate(s) (P) substituting for hydroxyl group(s) at specific positions, and omitting pyrophosphates a possible 63 stereochemical unique forms can exist. The standard nomenclature of InsPs ('Numbering of atoms in *myo*-inositol. Recommendations 1988. Nomenclature Committee of the International Union of Biochemistry', 1989) uses D-numbering of carbons and describes both the position of the phosphate groups and the number of phosphates, for example, *myo*-inositol 1,4,5-trisphosphate (or Ins(1,4,5)P₃) has D1, D4 and D5 carbons phosphorylated. Agranoff's turtle mnemonic (Agranoff, 1978) provides a useful visual aid (Figure 1.1b): where the turtle's front right limb is carbon-1, the turtle's head is carbon-2 and its remaining appendages are counted anti-clockwise; and a line of symmetry can be drawn from head to tail demonstrating enantiomer pairs (e.g. Ins(1,4,5)P₃ and Ins(3,5,6)P₃).

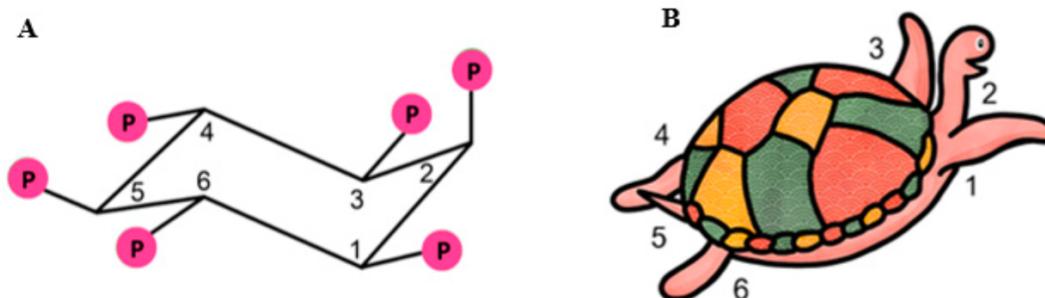


Figure 1.1 The structure of InsP_6 and Agranoff's turtle; showing (a) chair configuration of *myo*-inositol 1,2,3,4,5,6-hexakisphosphate, the most thermodynamically stable spatial arrangement of atoms, and (b) Agranoff's turtle with corresponding D-numbering of both (Agranoff, 1978). Figure from Chatree *et al.* (2020), permission to reuse the figure is not required as the article is published under Creative Common CC BY license.

Inositol pyrophosphates (also referred to as diphosphoinositol polyphosphates, PP-InsPs) are inositol polyphosphates which contain one or more highly energetic diester-phosphates and have similarly descriptive naming conventions, wherein the position of the pyrophosphate group is stated upfront, and the remaining positions of phosphates may be described similar to conventional InsPs. For example, 5-diphospho-*myo*-inositol 1,2,3,4,6-pentakisphosphate which has a pyrophosphate group on the D5 carbon and monoester-phosphates on D1, D2, D3, D4 and D6 carbons. The shorthand for this particular molecule may be noted as 5PP-Ins(1,2,3,4,6) P_5 , 5PP-Ins P_5 or 5-Ins P_7 , in this work the latter notation will be used. The exception to this being 5PP-Ins P_4 which refers to 5-diphospho-*myo*-inositol 1,3,4,6-tetrakisphosphate which contains six P groups but should not be confused with inositol hexakisphosphate i.e. InsP_6 (Shears, 2015).

1.2 Biosynthetic pathways of InsPs and PP-InsPs in plants

There is consensus of two divergent metabolic paths by which InsP_6 (and PP-InsPs) is/are synthesized in plants, the lipid dependent and the lipid independent pathway (Brearley and Hanke, 1996; Stevenson-Paulik *et al.*, 2005; Kuo *et al.*, 2018) [Figure 1.2]. The starting point for either pathway is the generation of *myo*-

inositol, the only inositol stereoisomer synthesised *de novo* in biological systems. The hexose sugar, glucose 6-phosphate, is the sole source of inositol in plants (Loewus and Murthy, 2000). This substrate is irreversibly isomerized into Ins(3)P₁ by *myo*-inositol-3-phosphate synthase (MIPS), a highly conserved enzyme throughout eukaryotes (Eisenberg *et al.*, 1964; Sherman *et al.*, 1969; Barnett *et al.*, 1973). There are three genes encoding MIPS proteins in Arabidopsis, with divergent functions (Donahue *et al.*, 2010). MIPS1 is expressed in all tissues and is contained within the cell cytoplasm (Donahue *et al.*, 2010). The InsP₁ product may then be dephosphorylated by inositol monophosphatase (IMP) to produce free D-*myo*-inositol (Loewus and Loewus, 1982). *myo*-Ins is an important cell metabolite which feeds into many metabolic pathways such as lipid signalling, stress response, cell wall biosynthesis and auxin perception (Loewus and Murthy, 2000). Free *myo*-Ins can also be converted back to Ins(3)P₁ by *myo*-inositol kinase (MIK), which is abundant in Arabidopsis and common bean (*Phaseolus vulgaris*) during seed development (Fileppi *et al.*, 2010; Kim and Tai, 2011).

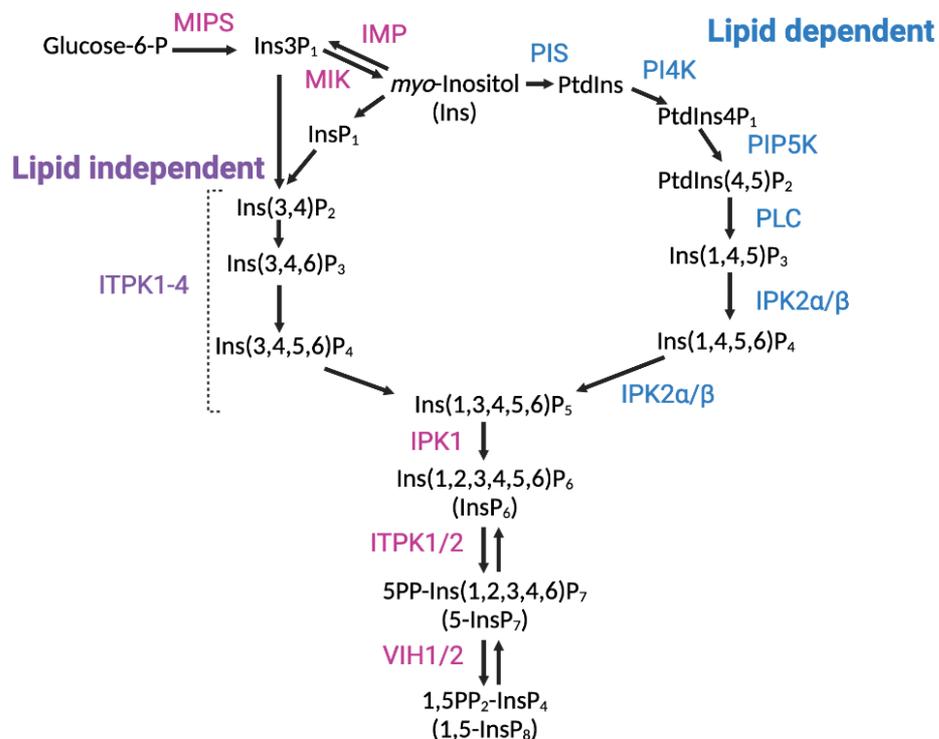


Figure 1.2 Biosynthesis of InsPs and PP-InsPs in plants. Highlighting enzymes involved in lipid independent pathway (purple), exclusive to lipid dependent pathway (blue), and enzymes common in both pathways (pink); the dashed line indicates steps not fully resolved. Figure adapted from Wang *et al.* (2021), created with Biorender.com.

The lipid dependent synthesis of InsP₆ was first described in budding yeast, *Saccharomyces cerevisiae* (York *et al.*, 1999), and has been broadly accepted as a pathway ubiquitous in all eukaryotes. Three enzymes are essential for this pathway: phospholipase C (PLC), inositol polyphosphate kinase 1 (IPK1) an InsP₅ 2-kinase, and inositol polyphosphate kinase 2 (IPK2 or IPMK) a multi-kinase (Odom *et al.*, 2000). As its name suggests, this metabolic pathway has cross over with inositol phospholipid signalling. Inositol phospholipids are generated from free *myo*-Ins which is converted to phosphatidylinositol (PtdIns) by phosphatidylinositol synthase (PIS) (Paulus and Kennedy, 1960), and then is sequentially phosphorylated at the D4 and D5 positions of the inositol ring by their respective inositol phosphatidyl kinases [phosphatidylinositol 4-kinase (PI4K) and phosphatidylinositol-4-phosphate 5-kinase (PIP5K)] (Balla, 2013). Phosphatidylinositol 4,5-bisphosphate (or PtdIns(4,5)P₂) is hydrolysed by PLC to diacylglycerol and the well characterised second messenger Ins(1,4,5)P₃ (Berridge, 1987). The inositol polyphosphate intermediates between Ins(1,4,5)P₃ and InsP₆ were determined via pulse-chase radiolabelling as Ins(1,4,5,6)P₄ and Ins(1,3,4,5,6)P₅ and inositol phosphate profiling with InsP₄ and InsP₅ standards (York *et al.*, 1999). In *Arabidopsis*, it is reported that these reactions are catalysed by two isomers of inositol polyphosphate 6-/3-kinase (IPK2 α and IPK2 β) (Stevenson-Paulik *et al.*, 2002, 2005).

The alternative lipid independent pathway was first described in the slime mould *Dictyostelium* (Stephens and Irvine, 1990), and duckweed (Brearley and Hanke, 1996), with common Ins(3,4,6)P₃ and Ins(1,3,4,5,6)P₅ intermediates. Although this pathway is not fully understood in plants, a family of inositol tris/tetraphosphate kinase (ITPK) enzymes are strongly implicated in several steps, not least because of the potent reduction in seed phytate observed in knock-outs of *AtITPK1* and *AtITPK4* (Kim and Tai, 2011). Humans possess a single ITPK ortholog and recent work (Desfougères *et al.*, 2019) indicates that the protein makes a significant, though poorly defined, contribution to InsP₆ synthesis in animals. In *Arabidopsis* there are four ITPKs (1-4) which have differential expression patterns and a predicted divergence of functions

(Sweetman *et al.*, 2006) with functional redundancy apparent between AtITPK1 and AtITPK2. Recently AtITPK4 was shown to be important in InsP₆ synthesis and normal profiles of InsP_{4s} in vegetative tissues, additionally it was demonstrated that AtITPK4 co-localize with AtITPK1 and AtIPK1 indicating that they are metabolizing the same pool of InsPs (Kuo *et al.*, 2018). As a point of convergence with the lipid dependent pathway, IPK1 catalyses the biosynthesis of InsP₆.

Plant PP-InsPs are synthesized by two diphosphoinositol pentakisphosphate kinase (PPIP5K) homologues, AtVIH1 and AtVIH2 (previously referred to as VIP2 and VIP1), and inositol 1,3,4-trisphosphate 5/6-kinase (ITPK) enzymes AtITPK1 and AtITPK2 (Desai *et al.* 2014; Laha *et al.* 2015, 2019). PPIP5K enzymes have dual 1-kinase/1-phosphatase activity, phosphorylating preferentially 5-InsP₇ to 1,5-InsP₈ over InsP₆ to 1-InsP₇; and hydrolysing 1,5-InsP₈ substrates to 1 or 5-InsP₇ (Zhu *et al.*, 2019). The dual functionality of PPIP5K enzymes has been suggested as an on/off signalling switch between InsP₇ and InsP₈ products (Adepoju *et al.*, 2019). The AtITPK1/2 enzymes (which phosphorylate InsP₃ and InsP₄ substrates with 1-kinase activity) are thought to generate the 5-InsP₇ substrate pool for InsP₈ synthesis, by phosphorylating InsP₆ (Laha *et al.*, 2019; Whitfield *et al.*, 2020).

1.3 Pi, InsP and PP-InsP cycling

Phytic acid (*myo*-inositol hexakisphosphate or InsP₆) is the most abundant InsP in eukaryotic cells (Sasakawa, 1995). In mature seeds, InsP₆ can contribute up to 80% of total phosphorus content (Raboy, 2009). This highly anionic molecule can prevent oxidative stress in seeds, which may increase germination capacity (Doria *et al.* 2009). During seed development, InsP₆ accumulates in protein globoids as phytate salts, in complex with mineral divalent cations such as calcium, manganese and zinc (For reviews: Loewus and Murthy, 2000; Raboy, 2003; Munnik and Nielson, 2011). There is variation in the content of phytic acid in different plants, and within the plant in specific tissues and cellular compartments throughout plant development. For example, in wheat or barley, phytate salts accumulate in the aleurone layer (Ockenden *et al.*, 2004; Regvar *et*

al., 2011), whereas phytate is most abundant in the embryo of *Arabidopsis* or maize seed (Otegui *et al.*, 2002; Pilu *et al.*, 2003). During germination, phytases break down the InsP_6 and release Pi, inositol and minerals into a nutritional pool (Loewus and Murthy, 2000).

In higher plants, Pi is the major form of phosphorus in vegetative tissues which is stored inside vacuoles (Yang *et al.*, 2017). Vacuoles, which in Pi sufficient conditions may contain 70-95% intracellular Pi (Bielecki, 1973), act as a buffer to Pi fluctuations in the cytoplasm, where tonoplast Pi exporters (such as the recently identified *Arabidopsis* PHT5 proteins) are responsive to Pi levels. Although present at much lower abundancies (than seeds), InsP_6 is synthesized by AtIPK1 which is constitutively expressed in roots, stem and leaves in vegetative cells (Phillippy *et al.*, 1994; Stevenson-Paulik *et al.*, 2005; Sweetman *et al.*, 2006). In alternative roles to phosphorus storage, InsP_6 has been implicated in 'signalling pathways' facilitative of e.g., guard cell function (Lemtiri-Chlieh *et al.*, 2000) and basal pathogen resistance (Murphy *et al.*, 2008), although direct attribution of roles to InsPs is difficult as InsP_6 is precursor to lower and higher InsP molecules.

The abundancies of PP- InsPs in plant cells are low, contributing to difficulties in PP- InsP detection and identification (Williams *et al.*, 2015). *Arabidopsis* seeds contain 2% PP- InsPs out of total inositol phosphate pool (with 1.33% InsP_7 and 0.24% InsP_8) and vegetative tissues including seedlings (0.64% InsP_7 and 0.14% InsP_8) vs. mature leaves (1% InsP_7), despite levels of InsP_6 (presumed precursor) being 2 orders more abundant in seeds vs. vegetative tissues (Desai *et al.*, 2014; Williams *et al.*, 2015). Methods distinguishing InsP_7 and InsP_8 isomers from plants have recently been published (Qiu *et al.*, 2020; Whitfield *et al.*, 2020; Riemer *et al.*, 2021), confirming the presence of specific PP- InsPs : 1- InsP_7 , 5- InsP_7 and 1,5- InsP_8 in plants. As cognate ligands of a family of SPX-domain-containing proteins, PP- InsPs , generated in conditions of phosphate excess, interfere with binding of SPX1 to the transcription factor PHR1 which binds to the P1BS promoter element of phosphate-starvation-response (PSR) or phosphate-

starvation-inducible (PSI) genes (Wild *et al.*, 2016; Jung *et al.*, 2018). 1,5-InsP₈, in particular, has been reported to regulate oligomerization and hence the promoter binding capability of PHRs through interaction of 1,5-InsP₈ with its SPX receptors (Ried *et al.*, 2021). Interestingly, however, the Arabidopsis *itpk4* mutation shows substantial decreases of InsP_{6/7/8}, just like *Atipkl* and *Atitpk1*, but without influence on Pi content and without constitutive activation of PSR genes under Pi-replete conditions (Figure 1.3; Kuo *et al.*, 2018; Wang *et al.*, 2021). It is also plausible that 1,5-InsP₈ interaction with SPX protein(s) is subject to competition by inositol phosphates, such as InsP₆, present at levels 2 orders of magnitude greater than 1,5-InsP₈ itself (Wang *et al.*, 2021); or that other InsP-dependent or -independent mechanisms impact SPX function.

1.4 Roles in abiotic stress response

Inositol phosphates are involved in a diverse range of signalling pathways in plant responses to abiotic stress, such as drought, salt, and cold (Jia *et al.*, 2019). Although detection and identification of specific InsP molecules within subcellular compartments remains elusive with current technologies, the presence or absence of specific InsPs is largely inferred by the localization of the enzymes involved in their synthesis and degradation, i.e., inositol phosphate kinases (IPKs) and phosphatases. For example, inositol polyphosphate 5-phosphatases (5PTases) which dephosphorylate Ins(1,4,5)P₃ which were found to be critical in the drought response pathway.

In guard cells, stomatal closure for water conservation is mediated through abscisic acid (ABA) drought hormone and calcium-mediated signalling (Schroeder *et al.*, 2001). Despite the importance of Ins(1,4,5)P₃ for calcium mobilisation from intracellular stores in animal cells (Berridge, 1993), in plants an involvement of this inositol phosphate molecule is not overwhelmingly confirmed. Ins(1,4,5)P₃ has been found to accumulate upon drought stress in relevant cells, in correlation to calcium levels increase (Khodakovskaya *et al.*, 2010). However, Ins(1,4,5)P₃ is not easily discriminated from other InsP₃ isomers in plants, plants do not possess G-protein-coupled receptors, nor PLC β orthologs

and importantly no InsP₃ receptor has been identified to connect a signal pathway (Krinke *et al.*, 2007). It has been hypothesised that InsP₃ accumulates only as a precursor to InsP₆ which has a similar positive correlation with ABA/calcium mediated response. InsP₆ is reported to release calcium from endoplasmic reticulum stores (Lemtiri-Chlieh *et al.*, 2003), however the theory suffers the same mystery of no known InsP₆ receptor for Ca²⁺-mediated signal transduction. Nevertheless, the family of higher InsP binding proteins, now extends from auxin, TIR and jasmonate, COII, co-receptor complexes (reviewed in Miyakawa and Tanokura 2017) to SPX family proteins (Wang *et al.*, 2021).

1.5 Modulation of signalling pathways through InsP binding

Apart from the enzymes involved with InsP biosynthesis (and degradation), InsP molecules have been found as structural co-factors in protein complexes where they have been described as a functional molecular glue. As highly electrostatic molecules, InsPs may increase binding affinity for the complex formation of two protein interaction partners (Chatree *et al.*, 2020). This function has been described for two plant F-box proteins TIR1 [Transport inhibitor response 1] (Tan *et al.*, 2007) and COII [Coronatine insensitive 1] (Sheard *et al.*, 2010; Mosblech *et al.*, 2011), which have been shown to bind specific InsP ligands discriminately/with different affinities. TIR1 is an auxin receptor involved in plant growth, development and gravitropism; whilst COII is crucial in jasmonate-mediated plant defence against herbivory. There is functional and structural similarity between these proteins and similar InsP binding sites within leucine rich repeat regions (Sheard *et al.*, 2010). In both cases, the InsP modulated hormone receptor complexes mediate degradation of target repressors in their respective pathways, through the activation of SCF (Skp1/Cdc53/E-box protein) or ASK (Arabidopsis Skp1-like) ubiquitin-ligase complexes (Xu *et al.*, 2002). The first descriptions of these InsP ligand-receptor complexes, found InsP₆ the co-factor of TIR1 (Tan *et al.*, 2007) and InsP₅ (2'OH) the proposed ligand of COII (Sheard *et al.*, 2010; Mosblech *et al.*, 2011) from analyses of X-ray crystallographic data and mass spectrometry. More recent studies have implicated PP-InsPs as physiological ligands of these proteins (Laha *et al.*, 2015, 2020).

Independent of Pi homeostasis, AtIPK1 and AtITPK1 have been shown to affect auxin signalling. In particular *Atitpk1* mutants showed phenotypes of altered auxin perception suggesting a role for 5-InsP₇ in this pathway (Laha *et al.*, 2020). Within this study, fluorescently tagged AtITPK1 was also shown to directly interact with TIR1 in co-immunoprecipitation and bimolecular fluorescence complementation assays *in planta* (Laha *et al.*, 2020). The authors propose a channel is formed between AtITPK1 and AtTIR1 to directly shuttle the physiological InsP product/co-ligand in order to promote auxin detection of the TIR1-ASK1 receptor complex and facilitate auxin signalling (Laha *et al.*, 2020).

In the case of COII, 1,5-InsP₈ has been suggested to modulate COII interaction with target JAZ (jasmonate ZIM domain) transcription repressors (Laha *et al.*, 2015). Here, the authors show that *vih* plant mutants [now known to be deficient in 1,5-InsP₈ (Dong *et al.*, 2019)] have impaired defence gene expression when challenged with insect and fungal pests (Laha *et al.*, 2015). The authors were able to demonstrate that radiolabelled InsP₅, InsP₆ and InsP₇, of unknown specific activity, extracted from plant tissues bound to the COI-ASK-JAZ1 complex and that InsP₆, 1-InsP₇, 4-InsP₇, 5-InsP₇ and 6-InsP₇ could all displace [³H]InsP₆ from the COII-ASK1-JAZ1 complex (Laha *et al.*, 2015). Additionally *in silico* modelling was used to assess the different InsP ligands (Laha *et al.*, 2015); thus demonstrating PP-InsPs as favourable ligands to the co-receptor complex.

The SPX (*Sygl/Pho81/Xpr1*) proteins (Dong *et al.*, 2019; Wang, *et al.*, 2021) represent a family of proteins conserved across plants, yeast and animals that are intimately associated with phosphate homeostasis. The family is expanded in plants and upon PP-InsP binding, one member, SPX1, has been shown to bind to PHR (phosphate starvation response) proteins, to regulate the expression of phosphate starvation response genes (Puga *et al.*, 2017; Zhu *et al.*, 2019). Through a process of considering many InsP ligands (see Chapter 4), 1,5-InsP₈ is thought to be the AtSPX1 physiological ligand and indicative of cellular Pi status (Dong *et al.*, 2019).

The growing number of cognate binding proteins identified for inositol phosphates in proteomic studies (Furkert *et al.*, 2021) make it highly likely that higher InsPs and PP-InsPs will yet prove to have multiple targets in diverse signalling processes in plants, regulated by the competition between different InsP and PP-InsPs in, as yet, unidentified subcellular locations.

1.6 Regulation of nuclear processes

InsP involvement in nuclear roles have been of keen interest, as studies have emerged of InsP-mediated mRNA export, transcription regulation, DNA repair and chromatin regulation described in yeast, plants and animal cells (Tsui and York, 2010; Lee *et al.*, 2015). Furthermore, inositol phosphate kinases such as ITPKs, IPMKs (IPK2s) and IPK1 are widely reported to participate in nuclear processes, localizing through undefined mechanisms to both the nucleus and the cytosol, while their InsP and PP-InsP products are reported cognate partners of the hormone receptor and other protein (SPX)-transcription factor complexes mentioned above.

In *Arabidopsis*, two ScIPK2 homologues have been cloned and characterized: AtIPK2 α and AtIPK2 β , where GFP-localization studies have demonstrated nuclear accumulation. The yeast multikinase ScIPK2 is a component of the ArgR-Mcm1 transcription complex involved in regulating arginine metabolism (Najet *et al.*, 2000). Various phenotypes of ScIPK2 mutants have been complemented with orthologs from diverse orders, and much controversy has ensued over catalytic requirement for this. Nevertheless, AtIPK2 β and other plant orthologs are sufficient to complement growth phenotypes of ScIPK2 mutants and in the case of a potato ortholog experimental verification of restored arginine metabolism/regulation to wildtype phenotype has been provided (Caddick *et al.*, 2007).

Atipk1-1 mutants have altered transcription profiles including phosphate starvation response (PSR) genes (Kuo *et al.*, 2014); as this mutant has a different

InsP profile compared to wt, particularly lower levels of InsP₆ and PP-InsPs, this may be explained by PSR mis-regulation due to low 1,5-InsP₈ and/or competition from other InsPs and PP-InsPs that are metabolic products of reversible kinase couples (Whitfield *et al.*, 2020). It is perhaps not surprising, in light of the explosion of interest in PP-InsPs, effectively derived from Ins(1,3,4,5,6)P₅ as the precursor of InsP₆, that the *Atipkl-1* mutant has multiple phenotypes, extending to altered chromatin composition via lower ARP6 mediated deposition H2A.Z (Kuo *et al.*, 2014). ARP6 is a component of the SWRI chromatin remodelling complex which is highly conserved in eukaryotes (Smith *et al.*, 2010).

The *Atipkl-1* mutant is, like its yeast counterpart (York *et al.*, 1999), also defective in export of messenger RNA from the nucleus. mRNA export is mediated by the nuclear pore complex through involvement of the protein Gle1. There is high conservation of function between Gle1 proteins in plants and yeast, in which it has been demonstrated that InsP₆ is essential in mediating the interaction between Gle1 with the DEAD-box protein [Dbp5] (Alcázar-Román *et al.*, 2006). The *Atipkl-1* mutant accumulates mRNA in the nucleus but can be rescued by expression in the *Atipkl-1* background of a modified Gle1 protein, with enhanced affinity for the low levels of InsP₆ found in *Atipkl-1* (Lee *et al.*, 2015).

1.7 Aims of project

As a protein with a zinc finger, a structural feature commonly associated with DNA binding, AtIPK1 will be investigated for DNA binding activity using a suite of methods including EMSAs, polarisation and FLIM-FRET (Chapter 2). Attempts to investigate AtIPK1-DNA interaction using X-ray crystallography could offer an insight into potential binding modes (Chapter 3). Methods to determine specific binding will be utilised should experiments confirm AtIPK1 as a DNA binding protein. Furthermore, inhibition investigations will be carried out to determine whether InsP ligands and small molecular inhibitors can alter AtIPK1 binding affinity to DNA. This will begin to elucidate whether AtIPK1's involvement in Pi homeostasis involves direct interaction with DNA, or an exclusively catalytic contribution thereto.

Similarly, Arabidopsis SPX domain proteins, involved in phosphate homeostasis and phosphate starvation response signalling cascade, will be cloned, expressed and characterised through fluorescence polarisation studies with fluorescein-tagged inositol phosphate probes (Chapter 4). Again, crystallography, which has been essential to define InsP binding sites in diverse proteins/cellular contexts, will be employed. The aim of the combination of the approaches is to understand InsP/PP-InsP binding affinities and how InsP/PP-InsP-SPX interactions could contribute to phosphate homeostasis. Determination of these preferences will provide a basis for investigations into how interaction of SPX domain proteins with phosphate starvation response transcription factors could be modulated by levels of inositol phosphate or pyrophosphates.

2 AtIPK1 interacts with DNA *in vitro*

2.1 Introduction

Inositol signalling pathways help co-ordinate several cellular processes (see details in Chapter 1). In recent years, roles have been given to inositol polyphosphates (InsPs) as regulators of nuclear processes (York, 2006; Seeds *et al.*, 2007; Monserrate and York, 2010; Tsui and York, 2010). For example, phytic acid or InsP₆ has been implicated in regulation of mRNA transport in yeast (York *et al.*, 1999), whereby genes that rescued a synthetic lesion in mRNA export were found to fall in three complementation groups, namely: PLC, IPK2 (also known as IPMK), and IPK1. InsP₆ is reported to control apoptosis of mammalian cancer cells (Liu *et al.*, 2015), and in plants is involved in chromatin regulation and other developmental processes (Kuo *et al.*, 2014) including those controlled by plant hormones such as auxin (Tan *et al.*, 2007) and jasmonate (Sheard *et al.*, 2010) that have nuclear receptors, TIR1 and COII, respectively. These functions have not been solely ascribed to single InsP species, since further InsP metabolism or other enzymes responsible for InsP turnover could be involved. Ubiquitously, InsP₆ synthesis is determined by the activity of inositol polyphosphate kinase-1 (IPK1 or inositol pentakisphosphate 2-kinase) which specifically phosphorylates the axial 2-hydroxyl of Ins(1,3,4,5,6)P₅ (York *et al.*, 1999; Verbsky *et al.*, 2002; Stevenson-Paulik *et al.*, 2005). The Arabidopsis InsP₅ 2-kinase gene product, hereafter referred to as AtIPK1, has been shown to compartmentalize to both the cytoplasm and cell nucleus of intact plant cells (Whitfield, 2013; Kuo *et al.*, 2014; Adepoju *et al.*, 2019), as has the human enzyme (Brehm *et al.*, 2007). In addition to this, unexpected zinc binding of AtIPK1 discovered through crystallographic approaches raises the possibility that this enzyme could have a role in novel nuclear functionality and specifically, direct DNA binding activity.

2.1.1 General characterization of AtIPK1

AtIPK1 was first cloned and characterised by two groups (Stevenson-Paulik *et al.*, 2005; Sweetman *et al.*, 2006), after a very comprehensive biochemical characterization of a soybean enzyme with the same activity (Phillippy *et al.*, 1994). The InsP₅ 2-kinase gene was identified using IPK1 sequence similarity

searches of the complete Arabidopsis genome; the Arabidopsis protein has low amino acid sequence identity to ScIPK1 [17.8%], but is more similar to hIPK1 [27.4%] (Sweetman *et al.*, 2006). Transcriptomic studies using RT-PCR and mRNA *in situ* hybridization were reported, showing that AtIPK1 is highly expressed in leaves and more weakly expressed in siliques and flowers (Stevenson-Paulik *et al.*, 2005). Purified AtIPK1 was used in kinase assays performed with different InsP substrates and analysed using HPLC to determine enzyme specificity and identify reaction products (Sweetman *et al.*, 2006). The InsP₅ 2-kinase activity was confirmed as turnover of InsP₅ to InsP₆; but additional activity of Ins(1,3,4,6)P₄ and Ins(1,4,5,6)P₄ to InsP₅ products were observed, albeit at a reduced efficiency (Stevenson-Paulik *et al.*, 2005). The 1:1 stoichiometry ratio of the InsP₅ and ATP substrates to InsP₆ and ADP products conversion was proposed and Michaelis-Menten kinetic parameters (V_{max} 35 nmol.min⁻¹.mg⁻¹ and K_m 22 μ M at 0.4 mM ATP) were obtained (Sweetman *et al.*, 2006). Substrate discrimination and kinase activation were further examined by X-ray crystallography, where the AtIPK1 structure was solved with InsP and nucleotide ligands; revealing the extensive contacts within the enzyme active site to orientate the axial 2-OH of InsP₅ for phosphorylation and the presence of metal ions, magnesium and zinc (González *et al.*, 2010). Interestingly, the non-catalytic zinc ion was coordinated by a zinc-binding motif consisting of residues H320, C330, C333 and H346 distal to the active site. These residues are highly conserved in plant IPK1 enzymes and assays suggested they are critical to protein folding (González *et al.*, 2010). The structure was further elucidated using mutant enzymes, InsP ligands and mimics to reveal shifts in protein domains upon nucleotide and substrate binding (Baños-Sanz *et al.*, 2012a; Gosein and Miller, 2013).

The physiological role for AtIPK1 was first explored with the *Atipkl-1* T-DNA insertional mutant line exhibiting several phenotypes: blocked InsP₆ accumulation in seeds, excessive accumulation of Pi in leaves, leaf epinasty and elongated root hair growth (Stevenson-Paulik *et al.*, 2005). These observations suggest that AtIPK1 is important in Pi regulation for the plant. More recently, it

was shown that the *Atipkl-1* mutant, containing an insertion in the last *AtIPK1* exon, is partially active within the plant cell, and therefore cannot fully describe the null mutant (Kuo *et al.*, 2014). Consequently, two additional insertion mutant lines were explored, named *Atipkl-2* (insertion within the sixth intron) and *Atipkl-3* (insertion in the third intron), which both abolished *AtIPK1* transcription and were lethal by approximately 20 days after germination (Kuo *et al.*, 2014). This demonstrated that *AtIPK1* is essential for later plant development. Further investigation of the transcriptome and phosphate starvation response (PSR) showed that *AtIPK1* is involved in transcriptional regulation of PSR genes via H2A.Z-mediated chromatin remodelling (Kuo *et al.*, 2014). At the time, it was proposed that an InsP ligand was responsible for this regulation, however more recent work assigns this function to the inositol pyrophosphate products of InsP₆ phosphorylation (see Chapter 4).

Noting the zinc binding motif in the *AtIPK1* structure, the Brearley initially investigated whether *AtIPK1* could have a more direct nuclear role, by conducting simple protein-DNA binding experiments utilizing fluorescence polarization (FP) and electrophoretic mobility shift assays (EMSA). These assays provided the first evidence of *AtIPK1* DNA binding activity *in vitro* (Whitfield, 2013). By the addition of an Alexa Fluor® 488-tag to *AtIPK1*, tagged-*AtIPK1*/*AtIPK1* dimerization was observed in FP assays upon additions of unlabelled *AtIPK1* (Whitfield, unpublished). Dimerization of *AtIPK1* also appeared to be favoured, over a monomeric state, in the presence of DNA (Whitfield, unpublished). These initial studies gave an insight into the potential role of *AtIPK1* as a DNA binding protein, but further analysis was needed by using a variety of DNA-protein interaction methods discussed below to provide robust evidence of such a novel interaction.

2.1.2 Investigative tools for DNA-protein interactions

Molecular biologists have an impressive assortment of tools at hand to investigate protein-DNA interactions (Cai and Huang, 2012); it is common for multiple techniques to be used to validate results and compensate for limitations

of any single experiment. In this study to investigate AtIPK1 for potential interaction with DNA *in vitro*, technology readily accessible at the School of Biological Sciences, UEA, was utilized. Protein-DNA interactions were explored using fluorescence polarization (FP), fluorescence-based electrophoresis mobility shift assays (F-EMSA) and analytical ultracentrifugation (AUC). Towards confirming the physiological relevance of AtIPK1 interaction with DNA and with the availability of the confocal fluorescence microscope and plant growth rooms on site, fluorescence lifetime imaging microscopy Förster resonance energy transfer (FLIM-FRET) analysis was attempted *in planta*. Here, these techniques will be described.

2.1.2.1 Fluorescence polarization

Fluorescence polarization (FP) or fluorescence anisotropy assays are highly sensitive methods used for detection of protein-DNA and protein-protein interactions (Jameson and Ross, 2010). This technique was first demonstrated by Weber (1952) and since the mid-1980s FP has become widely applied for high-throughput screening in pharmacological studies, as fluorescent molecules have become more accessible (Jameson and Ross, 2010). In principle, an immobile fluorochrome under polarized excitation light emits light in the same plane of polarization, more precisely – with discrete magnetic and electrical vector. Mobility of the probe causes the orientation of the fluorochrome to partially or wholly become randomized within the timescale of excitation, and the emitted light is less polarized (Perrin, 1926). Hydrodynamic properties cause higher molecular weight species to have slower rotation and give greater polarization values than lower molecular weight species (Figure 2.1).

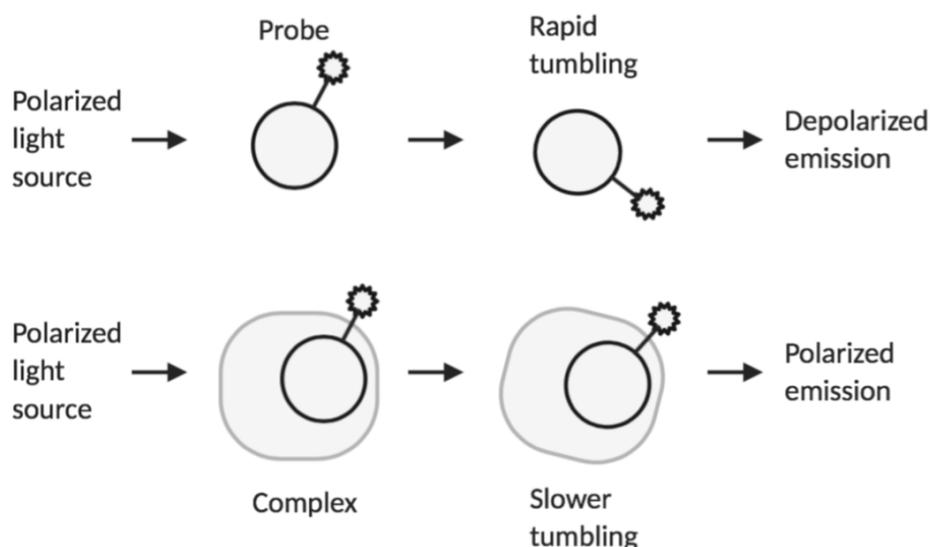


Figure 2.1 Diagrammatic representation of fluorescence polarization of unbound fluorescent probe and probe in complex. Figure adapted from Invitrogen 2006, created with Biorender.com.

A simple assay involves a mixture of fluorescently labelled molecule (such as a protein of interest, ligand or DNA) and different concentrations of a potential binding partner (such as protein or DNA) at equilibrium, and measuring polarization values (Jameson and Ross, 2010). Polarization is measured from the intensities of parallel and perpendicular with respect to the linear polarized excitation light source. Polarization values are calculated by the following equation:

$$P = \frac{(f_{\parallel} - f_{\perp})}{(f_{\parallel} + f_{\perp})}$$

where P = polarization, f_{\parallel} = fluorescence intensity parallel to excitation plane and f_{\perp} = fluorescence intensity perpendicular to excitation plane.

This method is a popular choice, as it directly measures the interaction of molecules in solution, requires no separation of the molecular species of interest and as an equilibrium experiment it can be used to obtain measurements of binding association and dissociation. However, as with other fluorescence

techniques, the presence of the fluorophore tag may interfere with the binding interaction that is being monitored.

2.1.2.2 Fluorescent electrophoretic mobility shift assay

Another traditional protein-DNA binding assay is the electrophoretic mobility shift assay (EMSA). This technique was first established in 1981 (Fried & Crothers, 1981), and a typical assay comprises of several binding reactions with a protein of interest and its potential DNA binding partner, with varying amounts of protein (Hellman and Fried, 2007). These reactions are then run on a NATIVE gel via electrophoresis, and the bands of DNA are visualized by radiolabelling or fluorescence labelling detection. An upwards shift of the DNA gel bands when the protein concentration is increased, suggest the formation of the protein-DNA complex retarding its mobility through the gel matrix (Figure 2.2). Several variations of this set-up can be arranged, where the concentration of DNA rather than protein is varied, the protein is monitored rather than the DNA, by the addition of different co-factors or carrier proteins to stabilize the formation of the complex or prevent loss of protein, putative inhibitors can also be monitored in this assay.

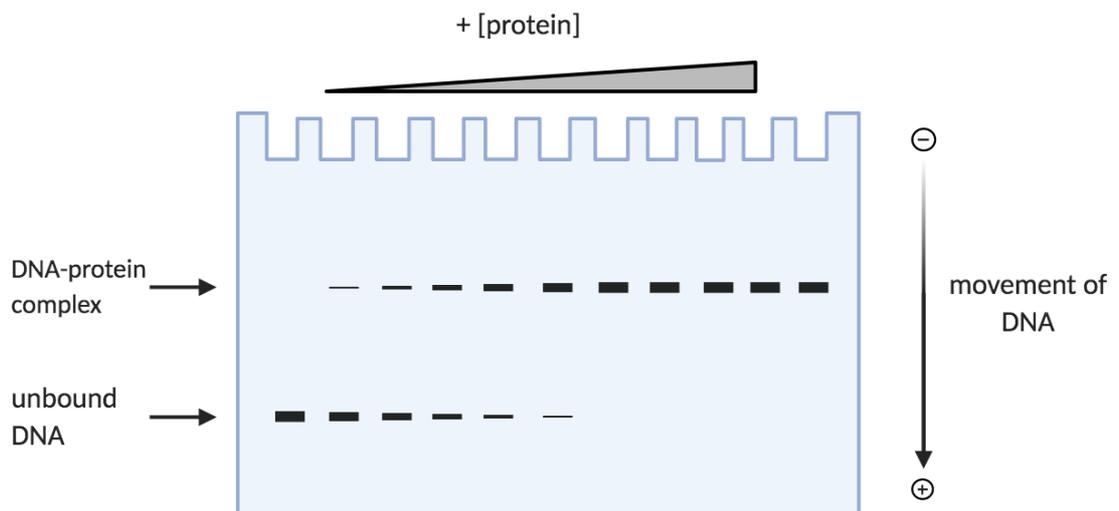


Figure 2.2 Diagrammatic representation of F-EMSA gel set-up where fluorescently labelled DNA is detected by fluorescence detectors. Image created with Biorender.com.

Although this method is quick to set-up and gel electrophoresis equipment is common in molecular biology labs, it can be difficult to obtain quantitative data for kinetic parameters. This can be somewhat circumvented by using fluorescently labelled DNA or proteins, and quantifying the fluorescence of each band (Hellman and Fried, 2007). However, the state of bound and unbound molecules is not at equilibrium at the point of imaging the gel, a feature of the movement and interaction of molecules with the gel matrix as the current is applied. Therefore, this tool though widely used is often used in conjunction with other protein-DNA assays.

2.1.2.3 Analytical ultracentrifugation sedimentation equilibrium

Almost 100 years after conception, analytical ultracentrifugation (AUC) analyses are versatile and useful for quantitatively characterizing proteins and protein interactions in solution (Cole *et al.*, 2008). AUC remains applicable in current research because of its foundation on fundamental physical principles (Svedberg and Pedersen, 1940) and high adaptability enabled by modern computational strategies for analysis (Zhao *et al.*, 2013). The premise of AUC is the analysis of concentration distribution of macromolecules in solution, subject to a large centrifugal force and monitoring sedimentation in real time via an optical detection system such as absorbance, interference fringes or fluorescence [Figure 2.3] (Cole *et al.*, 2008). From these studies, protein shape and size, biomolecular associations and conformational changes can be modelled from experimental data. Unfortunately, these methods are limited by the high cost of the required equipment (for example the ultracentrifuge, optics systems and specialized double-sector window cells) and technical expertise.

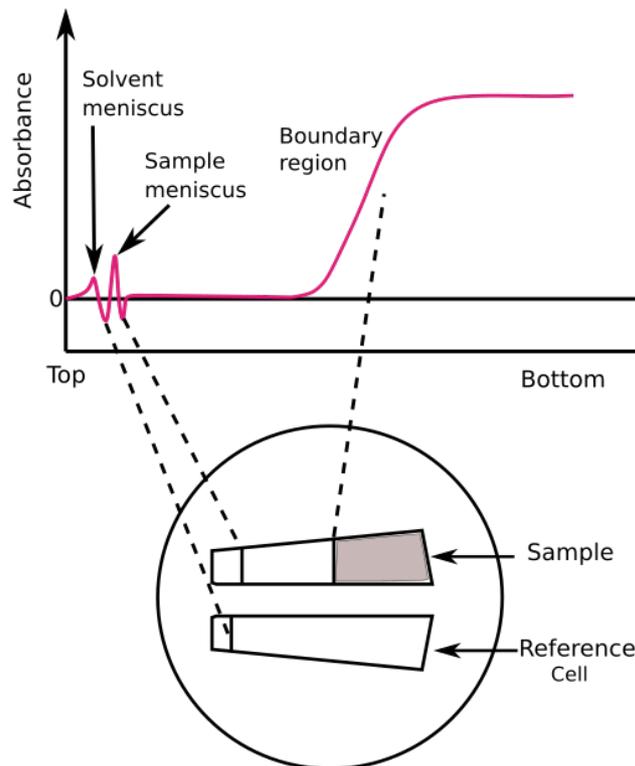


Figure 2.3 Representation of absorbance scan data from protein in solution (at A280) in an AUC experiment, with sample and reference in specialized cell with quartz windows. Adapted from Introduction to analytical ultracentrifugation, Beckman Coulter, Life sciences, Ralston.

There are two approaches of AUC, sedimentation velocity (SV) and sedimentation equilibrium (SE) experiments (Cole *et al.*, 2008). The former monitors the evolution of sedimentation in a time-dependent manner and uses hydrodynamic principles to answer questions of macromolecular shape, molar mass and size distribution (Howlett *et al.*, 2006; Cole *et al.*, 2008). Whereas, SE analyses are based on thermodynamic principles and can be used for: determining oligomeric state, protein molecular mass, non-ideality, protein-ligand interaction affinity and stoichiometry of interactions (Laue, 1995). In this study, AUC-SE analyses were chosen to explore the AtIPK1-DNA interaction.

As implied by its name, SE is an analysis of the final analyte distribution when equilibrium is reached. This is achieved when a macromolecular solution is centrifuged at a constant rotor velocity and a time-independent gradient is attained that reflects the transport of solute due to centrifugal force and opposing

back-diffusion (Cole *et al.*, 2008). A typical SE experiment requires presentation of high purity macromolecular samples and long experimental run times of > 24 h; therefore, proteins must stay stable throughout the duration of the experiment. If these conditions can be met, it can offer a valuable insight into the nature of interaction with DNA, for example quaternary structure or buffer conditions under which the interaction occurs.

2.1.2.4 FLIM-FRET

Förster resonance energy transfer (FRET) is used to investigate *in situ* protein intermolecular interactions (Blouin *et al.*, 2009). In the mid-1940s, Theodore Förster observed and described the eponymous process by which protein interactions between fluorescently labelled molecules may be observed (Förster, 1946). FRET occurs through the non-radiative energy transfer of an excited donor molecule to a ground state acceptor molecule by fluorescence, phosphorescence or chemiluminescence (Figure 2.4a), and therefore requires that the emission spectrum of the donor overlaps with absorbance spectrum of the acceptor (Figure 2.4b) i.e., a FRET pair.

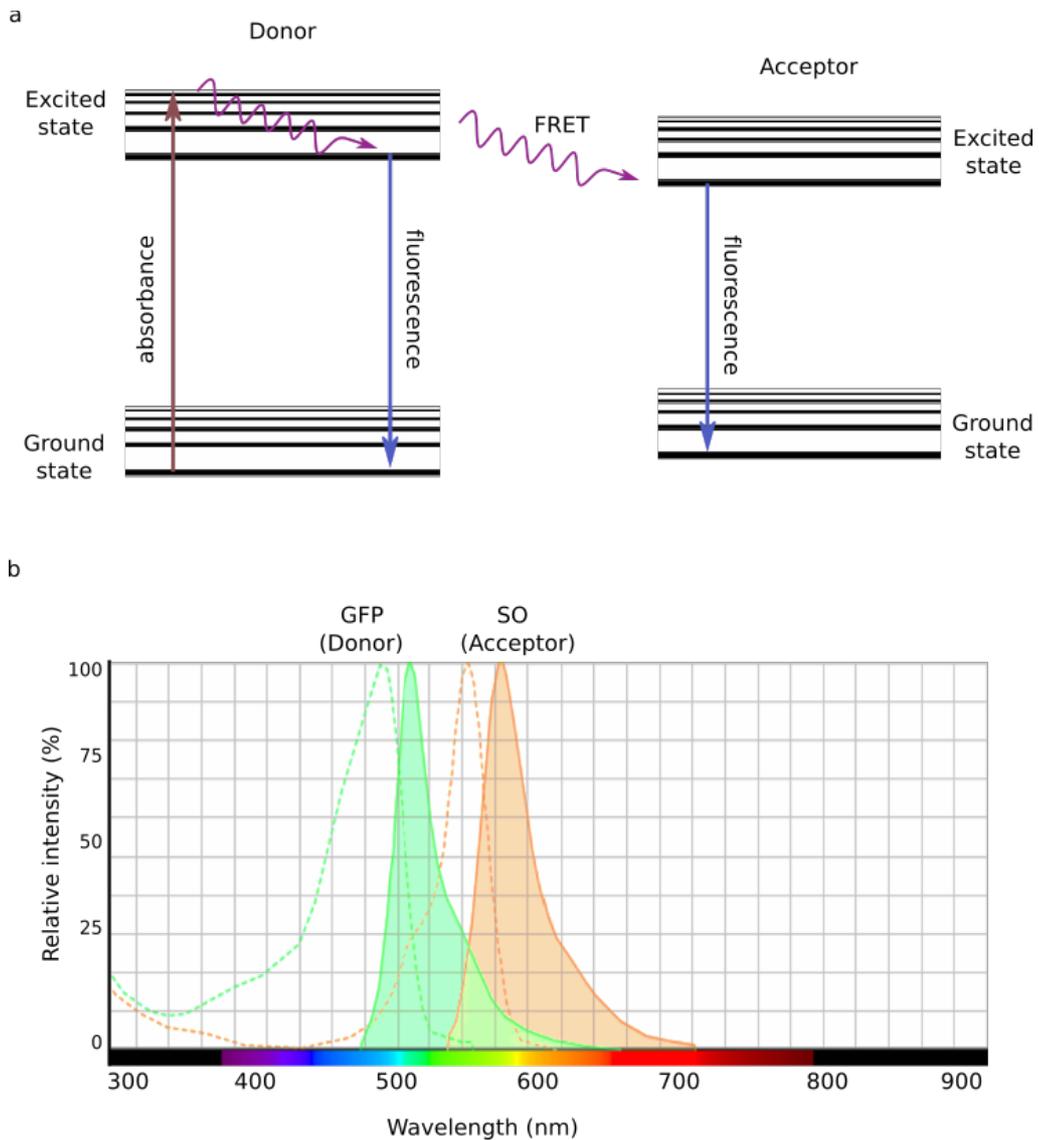


Figure 2.4 Simple illustration of FRET principles. (a) Jablonski plot of energy transfer, where straight horizontal lines show different energy states, vertical arrows show photon absorbance or emission, wavy arrows show non-radiative energy transfer. (b) Overlap of FRET-pair GFP (green) and SO (orange) emission (solid line, shaded) and absorbance spectra (dashed line).

FRET frequency is directly proportional to the distance between the FRET pair when in the correct orientation of dipole-dipole coupling, of which distances between donor and acceptor are detectable within 2 to 10 nm (Blouin *et al.*, 2009). At these short intermolecular distances, collisions occur between donor and acceptor molecules and FRET can be observed, indicating that the protein-interaction is physiological.

FRET may be detected using fluorescence intensity, spectrometry, or fluorescent lifetime imaging microscopy (FLIM). The latter of which has recently been applied to plant systems (Bücherl *et al.*, 2014), and was used in this current investigation. FLIM is an imaging technique that monitors the decay rate of individual ensembles of or single fluorophores from an excited state to ground state (Datta *et al.*, 2020) and is defined by the average time the donor fluorophore remains excited, τ (Figure 2.5). Time Correlated Single-Photon Counting (TCSPC) is often used to measure FLIM, it enables measurements of the time difference on nanosecond timescale between the excitation pulse and detection of the emitted photon. When FRET occurs, the donor emission is quenched and the lifetime decreases, because the timescale of the non-radiative process (typically $10^{-14} - 10^{-11}$ s) is shorter than that of the radiative process [$10^{-9} - 10^{-7}$ s] (Bajar *et al.*, 2016). Therefore, FRET efficiency can be calculated by comparing the average lifetimes of donor in the presence and absence of the acceptor.

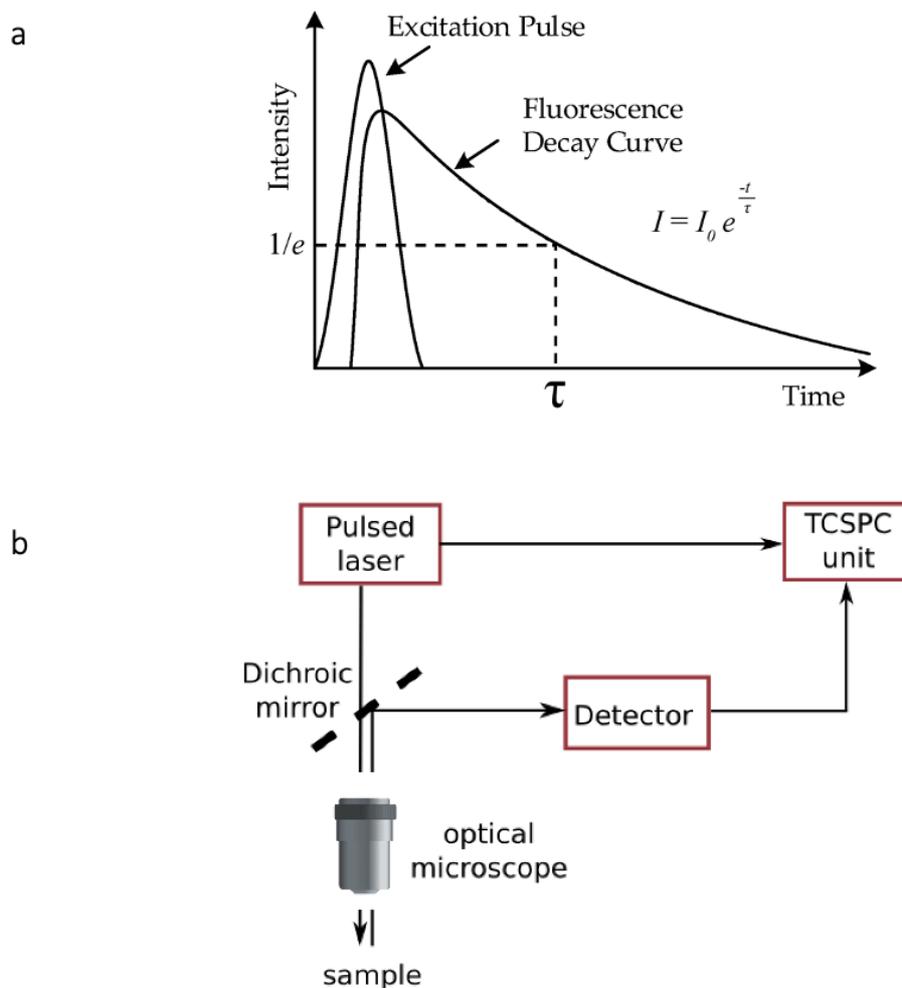


Figure 2.5 Fluorescent lifetime imaging microscopy. (a) Fluorescence lifetime decay curve (adapted from Wei *et al.* 2017), where the lifetime τ is the time required for fluorescence intensity (I) to decay to $1/e$ of its initial value I_0 . (b) Schematic of FLIM detection set-up.

In advantage over fluorescence intensity methods of FRET detection, FLIM measurements are independent of concentration of fluorophores (Bajar *et al.*, 2016), and therefore amenable to variations present in biological samples. However, to achieve a high signal to noise ratio, data acquisition per measurement can take minutes (Bajar *et al.*, 2016). As with other FRET techniques, expertise to obtain valid and replicable data in addition to access of expensive equipment such as a TCSPC unit, can hinder its use.

2.1.3 Aims of experiments

In this study, protein-DNA investigative tools were used to explore the AtIPK1-DNA interaction. AtIPK1 has previously been shown to bind DNA non-specifically *in vitro* using FP and EMSA (Whitfield, unpublished). Non-specific DNA binding is primarily facilitated by electrostatic interactions between charged molecules (von Hippel and Berg, 1986), therefore, to further investigate non-specific DNA binding of AtIPK1: (1) salt additions to the binding buffer were used to limit electrostatic interactions between protein and DNA; (2) different nucleotide oligomers were synthesized in order to examine sequence preferences.

FP was utilized to obtain quantitative binding data for points of comparison within the study, and in particular the dissociation constant, K_d , enables comparisons to other plant DNA binding proteins. Additionally, F-EMSA was used to visualize the formation of the AtIPK1-DNA complex, for validation.

Previous work has shown that AtIPK1 dimerization is favoured in the presence of DNA (Whitfield, unpublished). To examine the hypothesis that AtIPK1 binds to DNA as a dimer, AUC-SE was used to (1) infer the oligomeric state of AtIPK1 in the presence and absence of DNA, (2) evaluate the protein-DNA interaction without the use of fluorescent labelling.

Finally, FLIM-FRET was attempted to test whether the AtIPK1-DNA interaction is physiologically relevant, by calculating FRET efficiency of transiently expressed GFP labelled AtIPK1 and GFP, as a negative control.

2.2 Material and methods

2.2.1 Expression and purification of AtIPK1

2.2.1.1 Propagation of *pOPINF[T7lacO/AtIPK1]* in Stellar

Full length *AtIPK1* DNA in *pOPINF* was provided by Hayley Whitfield from Stellar stocks stored at -80 °C. The plasmid was propagated in Stellar cells (Takara bio) with 10 mL LB and 100 µg/mL ampicillin overnight at 37 °C, 200 rpm and isolated using the Wizard® *Plus* SV MiniPreps DNA Purification System (Promega) as per the manufacturer's instructions. Successful preparations were verified using OD readings with NanoDrop™ One (Thermofisher) at A260/A280 for quick estimates of DNA concentration and DNA gel electrophoresis for visual confirmation on agarose gels. Purified plasmids were stored at -20 °C.

2.2.1.2 DNA gel electrophoresis

To prepare for gel electrophoresis, 100 mL 1% agarose solution (in 1x TAE buffer: 40 mM Tris, 20 mM acetic acid, 1 mM EDTA) was heated in a 600 W microwave at full power for 2 min, until the agarose was fully melted. Once cooled to approximately 50 °C, ethidium bromide was added to a final concentration of 0.5 µg/mL and the solution was poured into a sealed gel tray to set at room temperature (rt). DNA samples or PCR products were loaded into wells with 5x loading dye (Invitrogen) alongside a 1 kb plus DNA ladder (Invitrogen) for size identification. Gels were run at 100 V for 45 min in 1x TAE buffer and imaged under UV light using the Bio-Rad ChemiDoc Gel Imaging System (Bio-Rad).

2.2.1.3 Heat shock transformations into competent cells

The *pOPINF[T7lacO/AtIPK1]* plasmid was transformed into competent *E. coli* strains including Stellar (Takara bio), NEB stable (NEB) and Rosetta (Novagen). Competent cells were thawed on ice and 100 µL aliquots were transferred into pre-chilled 1.5 mL microcentrifuge tubes and incubated with 10 ng plasmid DNA on ice for 30 min. Heat shock was then performed at 42 °C in a water bath for 45 s, before immediately resting on ice for 5 min. An aliquot of 1 mL LB was then added to each reaction and the mixture was incubated at 37 °C, 200 rpm for 1 h of recovery. Transformation reactions were then spread onto LB agar plates

containing appropriate antibiotics (100 µg/mL ampicillin) for selection and stored inverted at 37 °C overnight. Putative transformants were identified the next day by their presence as growing colonies.

2.2.1.4 Glycerol stocks for *pOPINF [T7lacO/AtIPK1]*

To make glycerol stocks, three colonies were picked from each transformation plate and grown in 10 mL LB media with selective antibiotics (100 µg/ml ampicillin) overnight at 37 °C, 180 rpm. Overnight cultures were spun down gently, media was tipped off and the cells were resuspended in the residual liquid. Approximately 40 µL of the cell suspension were spread onto each LB agar plate (or LBG agar plate for Rosetta cells) containing relevant antibiotics. Two plates were used to generate each stock. The plates were then inverted and stored in a 37 °C incubator overnight to allow bacteria to grow.

For each pair of plates, 1.5 mL of LB or LBG media were added to one culture plate. The cells were scraped into the fresh media using a glass spreader and the liquid was then transferred to the second plate where the cells were similarly scraped. The collected cells, approximately 1 mL, were transferred into sterile microcentrifuge tubes and 333 µL of 60% glycerol was added. The stocks were inverted for mixing before snap freezing in liquid nitrogen. Glycerol stocks were stored at -80 °C.

2.2.1.5 Large scale *AtIPK1* expression and cell harvest

Stocks of Rosetta cells with *pOPINF [T7lacO/AtIPK1]* were used to produce a subculture in 100 mL of LBG media, utilising glucose to repress the lac promoter thereby preventing expression prior to IPTG induction. This culture was grown with ampicillin (100 µg/mL) for selection overnight at 37 °C with shaking (180 rpm) for maximal growth, and subsequently was used to inoculate 100 mL starter cultures (LBG media containing 100 µg/mL ampicillin), which were grown overnight under the same conditions.

The following day, cultures were gently pelleted at 1700x g for 10 min at 4 °C and LBG discarded. Pellets were resuspended with a small amount of LB and combined so that 300 mL of starter culture would be used to inoculate 1 L of LB with 100 µg/mL ampicillin for large scale expression. The 1 L cultures were immediately induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and expressed overnight at 18 °C, 180 rpm.

Expression cultures were finally transferred into 1 L centrifuge pots on ice and spun at 5000 rpm in a Beckman Avanti J-20 with a JLA8.1000 rotor for 20 min. Pairs of pellets were combined by resuspending with 25 mL of lysis buffer (50 mM NaH₂PO₄ pH 7.5, 300 mM NaCl, 20 mM imidazole, 0.5% triton) before being transferred into 50 mL corning tubes and snap frozen with liquid nitrogen. To verify successful protein expression, samples of 100 µL were collected from cell suspensions for SDS-PAGE analysis, before storage at -80 °C or immediately beginning cell lysis and protein purification steps.

2.2.1.6 French pressure cell

Cell lysis was achieved through use of the French press at 1000 psi. Cells in resuspension buffer were thawed on ice and then the culture was passed through the prechilled cell in ~35 mL batches twice, whilst being kept on ice. Lysed cells were then transferred and balanced in centrifuge pots designed for the 45 Ti rotor. Lysates were spun in the Beckman Optima L-90K Ultracentrifuge at 4 °C at 40,000 rpm for 45 min, and the supernatant from each pot was collected and kept on ice for immediate use. Finally, a 20 µL sample was taken for SDS-PAGE analysis to test for soluble protein.

2.2.1.7 Purification of AtIPK1

The protocol for His-tagged AtIPK1 purification has been previously established (Sweetman *et al.*, 2006; González *et al.*, 2010; Whitfield, 2013). All purification steps described hereafter, except when stated otherwise, were conducted using the AKTA Pure system (GE Healthcare) at 4 °C ambient temperature, with the

Unicorn 7 software package for method design and data analysis (GE Healthcare).

Nickel affinity chromatography was performed with a 5 mL NiNTA HiLoad column (Qiagen) equilibrated with 2x column volumes (cv) of NiNTA buffer A (50 mM NaH₂PO₄ pH 7.5, 300 mM NaCl, 20 mM imidazole). The supernatant from the cell lysate was applied to the column at 0.5 mL/min, followed by a column wash with 6x cv NiNTA buffer A at 1 ml/min. An imidazole gradient to elute the protein was achieved with NiNTA buffer A and 0 to 100% NiNTA buffer B (50 mM NaH₂PO₄ pH 7.5, 300 mM NaCl, 500 mM imidazole) at 1 mL/min over 100 mL. Fractions were collected in 2 mL aliquots in a 96 well plate (GE Healthcare). A final elution step of 6x cv NiNTA buffer B ensured that any column bound proteins were removed. The UV trace was used to determine protein containing fractions which were subsequently sampled for verification by SDS-PAGE analysis. The verified fractions were then pooled together.

To remove nucleotide contamination of AtIPK1, an additional purification step using heparin chromatography was performed. The protein fractions pooled after nickel affinity chromatography were diluted 6x with 20 mM TrisHCl pH 7.5, in order to reduce salt content. The His-protein was then loaded at 0.2 mL/min onto a 1 mL Heparin HiTrap column (Qiagen) equilibrated with 10x cv heparin A buffer (20 mM TrisHCl pH 7.5, 50 mM NaCl). The bound protein was washed with 10x cv of heparin A buffer, and then an elution gradient was applied from 100% heparin A buffer to 100% heparin B buffer (20 mM TrisHCl pH 7.5, 1 M NaCl) at 0.5 mL/min over 100 mL. Eluted protein was collected in 2 mL fractions, and a final elution step of 10x cv heparin B buffer was observed. The UV trace determined protein-containing fractions were verified by SDS PAGE analysis, and consequently pooled together. The level of nucleotide contamination was observed using the nanodrop one (Thermofisher) A260/A280 ratio. An estimate of the amount of protein present (A280) was determined, where ProtParam (ExpASY) was used to calculate the extinction coefficient of AtIPK1 (0.915 M⁻¹cm⁻¹) to allow conversion to protein concentration.

The 6xHis tag was cleaved from AtIPK1 using HRC 3C protease (ThermoFisher Scientific Pierce) in recommended amounts from the manufacturer's instruction depending on the estimated protein content. Before the addition of enzyme, the His-IPK1 was concentrated to a smaller volume, ~ 3 mL, using a 50 mL Amicon® stirred cell, with 10 kDa mwco membrane filter (Ultracel), in order to improve enzyme efficiency. The reaction was incubated at 4 °C overnight.

The cleaved AtIPK1 was collected by loading the overnight reaction onto a pre-equilibrated (6x cv NiNTA buffer A) 5 mL NiNTA HiLoad column (Qiagen) at 1 mL/min. The column was washed with NiNTA buffer A at 1 mL/min, and the flow through was collected in 1 mL fractions. AtIPK1 containing fractions were identified by the UV trace and pooled together.

Lastly, size exclusion chromatography was used to collect purified, untagged AtIPK1, in a single conformation. The cleaved protein was concentrated to 2 mL using a 50 mL Amicon® stirred cell followed by a 10 mL Amicon®, with appropriately sized 10 kDa mwco membrane filters (Ultracel). This was applied manually with a 2 mL loop at 1 ml/min onto a Sepharose Superdex 75 16/600 column equilibrated with 2x cv of gel filtration buffer (20 mM Tris-HCl, pH 7.5, 200 mM NaCl). Isocratic elution was performed with 2x cv at 1 ml/min. The first 30 mL were not collected, whereas subsequent 2 mL fractions were collected in the 96 well plate. The protein fractions under a single UV peak were collected, pooled together, and concentrated using a Sartorius Polyethersulfone (PES) membrane 10 kDa mwco spin concentrator as per manual instruction to desired protein concentration determined by nanodrop readings. Purified protein that was not immediately used, was stored in 20 µL aliquots into PCR tubes and snap frozen in liquid nitrogen for storage in the -80 °C freezer.

2.2.1.8 SDS-PAGE analysis of cell pellets, supernatant and purified AtIPK1

Samples of expression cultures in suspension were centrifuged at high speed (13,000 rpm) for 5 min, the supernatant discarded and pellets resuspended with

1x SDS loading dye (50 mM Tris-HCl pH 6.8, 2% (w/v) SDS, 10% (v/v) glycerol, 100 mM DTT, 0.1% (w/v) bromophenol blue). Protein samples from soluble fractions were diluted appropriately with 4x SDS buffer. All samples were boiled for 5-10 min and either used immediately or stored at -20 °C.

SDS polyacrylamide gels, 12% resolving (12% acrylamide/bisacrylamide solution, 375 mM 1.5 M Tris HCl pH 8.8, 0.1% (w/v) SDS, 0.1% APS, 4-10 µL (0.04%-0.1%) TEMED) and 5% stacking (5% acrylamide/ bisacrylamide solution, 125 mM 1 M Tris HCl pH 6.8, 0.1% (w/v) SDS, 0.1% APS, 5 µL (0.1%) TEMED), were assembled with the Mini-PROTEAN 2 system (Bio-Rad) and loaded with 20 µL of sample or 10 µL precision plus protein standard blue ladder marker (Bio-Rad). A 1x SDS running solution (25 mM Tris base, 192 mM glycine, 0.1% SDS) was used for vertical gel electrophoresis at 100 V for 1 h. The gel was stained with 30 mL InstantBlue™ solution (Expedeon) and left agitating for 1 h before being destained with distilled water until the protein bands became clear. SDS-PAGE gels were imaged under epi-white illumination with the ChemiDoc Gel Imaging System (Bio-Rad).

2.2.2 Protein-DNA binding assays

2.2.2.1 Oligomer design and synthesis

Single stranded oligomers and modified oligomers (Table 2.1) were synthesised by Eurofins genomics. Equimolar complimentary ssDNA were added to the desired final concentration, with either Invitrogen 1x T4 DNA ligase buffer (50 mM Tris-HCl pH 7.6, 10 mM MgCl₂, 1 mM ATP) or annealing buffer (10 mM Tris-HCl pH 7.5, 50 mM NaCl, 1 mM EDTA). Annealing was achieved in 100 µL reaction volumes using a thermocycler for heating and cooling as follows: 95 °C for 2 min, then 96 cycles of a 1 °C decrease per 1 min and finally held at 4 °C. DNA was stored at -20 °C or used immediately as required. Annealing was verified using NATIVE-PAGE with 20% polyacrylamide gels (20% acrylamide/ bisacrylamide solution, 375 mM, 1.5 M Tris-HCl pH 8.8, 0.1% APS, 0.04% TEMED) and ethidium bromide staining for visualization under UV light on ChemiDoc Gel Imaging System (Bio-Rad).

Table 2-1 DNA oligomers synthesized by Eurofins.

Oligomer name	Modification	Sequence 5' → 3'
GC rich F	5'-FAM	CCCCGGGTACCGAGCTCGAATTCAGTGGCCGTC GTTTTAC
GC rich R		GTAAAACGACGGCCAGTGAATTCGAGCTCGGTA CCCCGGG
AT rich F	5'-FAM	CCCCGGCGAATTAATTCGAATTCAGTAAACGTA GTTTTAC
AT rich R		GTAAAACGACGGTTAGTGAATTCGAATTAATTC GCCCCGGG
12mer F	5'-FAM	CGAATTAATTCG
12mer		CGAATTAATTCG

2.2.2.2 Fluorescence polarization with AtIPK1

Fluorescence polarization (FP) experiments were performed in 100 μ L reaction volumes with 20 mM HEPES pH 7.5, 1 mM MgCl₂ buffer containing 2 nM fluorescent probe (for FAM-oligomers see Table 2.1; and 2-FAM-InsP₅ see Appendix 1) and increasing amounts of AtIPK1 or AtITPK4 negative control (0 to 10 μ M). The buffer was adjusted with the addition of NaCl (50, 100, 200 mM) to reduce non-specific electrostatic interactions between molecules. The reactions were incubated for 1 h on ice or at rt before quadruplicate replicates were dispensed in 20 μ L aliquots into a 384 well Corning low volume, non-binding, black, flat bottomed, polystyrene plates (Corning #3575). Polarization was measured at 25 °C using a Clariostar (BMG LabTech) plate reader with excitation at 485 \pm 5 nm and emission at 520 \pm 5 nm.

The raw data was used to calculate the mean and standard deviations of ≥ 3 separate measurements using Prism v6.0 (GraphPad). Non-linear regression analysis with variable slope model was used to produce standard curves and obtain EC50 measurements. Transformations of the data into anisotropy data and calculating fraction bound were made in Excel. Finally, Prism v6.0 (GraphPad) was used to perform a one site model total binding analysis using the

fraction bound data, and K_d was estimated. This approach makes the assumption that the extent of non-specific binding is proportional to the concentration of the probe.

2.2.2.3 F-EMSA

The fluorescence-based electrophoretic mobility shift assay (F-EMSA) has been described previously (Steiner and Pfannschmidt, 2009). Binding reactions were performed in 20 μ L reactions with 2.5 nM FAM-oligomer and different amounts of AtIPK1 (0-10 μ M) or AtITPK4 (0-10 μ M) negative control, in 20 mM HEPES pH 7.5, 1 mM $MgCl_2$, 50 mM NaCl and kept on ice for 1 h. Additional components to limit non-specific interactions and enzyme activity included the optional inclusions of BSA (0, 0.01, 0.05, 0.1 mg/mL), NaCl (0-150 mM) and EDTA (0, 1 mM). Also, to help with loading and tracking the progression of binding reaction through the gel, native loading dye (50 mM Tris-HCl pH 6.8, 0.01% (wv) bromophenol blue, 10% (v/v) glycerol) or 5% glycerol was used. These variations where used are described in the text.

The reaction products were separated on an 8% native polyacrylamide gel (8% acrylamide/bisacrylamide, 375 mM Tris pH 8.8, 0.1% APS, 0.05% TEMED) with 5% native stacking gel (5% acrylamide/bisacrylamide, 125 mM Tris pH 6.8, 0.1% APS, 0.05% TEMED) which were pre-run at 50 V for 2 h on ice. After loading the wells with 5 μ L of binding reactions, the gels were run at 50 V for a further 2-2.5 h on ice. The gel assemblies were then taken apart, gel plates removed and gels were dried and transported in cling wrap before being exposed to the pre-set Typhoon™ FLA 9500 system (GE Healthcare) parameters for FAM detection (excitation at 490 nm and emission 520 nm). To locate the proteins on the native gel, InstantBlue™ (Expedeon) was used to stain the gels for 1 h with 50 rpm agitation, followed by destaining for a minimum of 1 h, with 50 rpm. Gels were finally imaged under epi-white illumination with the ChemiDoc Gel Imaging System (Bio-Rad).

Gel images of fluorescence and protein detection were overlain in GIMP v2.8 and the binding data was quantified using Fiji (Schindelin *et al.*, 2012) gel analyser, where each measurement reported represents the mean of four independent measurements \pm the standard deviation (SD).

2.2.2.4 Analytical ultracentrifuge sedimentation equilibrium

Analytical ultracentrifuge sedimentation equilibrium (AUC-SE) experiments were performed using a Beckman Optima XL-1 analytical ultracentrifuge with scanning absorbance optics and an An50 Ti rotor. Initially, three AtIPK1 samples were diluted in buffer (20 mM HEPES pH 7.5, 1 mM MgCl₂ and 50 mM NaCl) to a final concentration between 5 and 15 μ M. Samples of 105 μ L were loaded into charcoal-filled Epon double sector cells fitted with quartz windows. Reference sectors were filled with 120 μ L of corresponding buffer and centrifugation was performed at 20°C. Absorbance profiles at 280 nm (A₂₈₀) were collected by taking 10-15 scans every 5 h at speeds of 10,000 and 11,000 rpm for three days in total. SE was established when profiles were observed to be constant between two timepoints. Sample degradation was concurrently monitored over this time at rt, using the same protein stocks; these were prepared for SDS-PAGE analysis (as described in 2.1.1.8.).

To find the most suitable conditions for further experiments, AtIPK1 sedimentation profiles of five buffers (detailed in Table 2.2) were monitored over the course of 48 h. SE was performed as described above, wherein 10 μ M AtIPK1 was used in each buffer condition, and the rotor speed was constant at 10,000 or 16,000 rpm. In all subsequent SE experiments 20 mM HEPES pH 7.5, 1 mM MgCl₂, 50 mM NaCl and 0.5 mM DTT was used to dilute samples.

Table 2-2 Buffers used for optimization trial in SE experiments with 10 μ M AtIPK1

Buffer no.	Buffer composition
1	20 mM HEPES pH 7.5, 50 mM NaCl
2	20 mM Tris pH 8.0, 50 mM NaCl
3	20 mM HEPES pH 7.5, 50 mM NaCl, 1 mM MgCl ₂
4	20 mM HEPES pH 7.5, 50 mM NaCl, 1 mM MgCl ₂ , 2 mM DTT
5	20 mM HEPES pH 7.5, 50 mM NaCl, 1 mM MgCl ₂ , 1 mM TCEP

DNA samples (of annealed 5'-CGAATTAATTTCG-3') were prepared to final concentrations between 2 and 4 μ M, and sedimentation was monitored at A260. Different DNA-protein micromolecular ratios in two independent experiments (experiment 1 at DNA:protein 2:0, 2:4, 2:6, 1:2, 1:4 μ M; and experiment 2 at DNA:protein 2:0, 1:1, 2:2, 3:2, 4:2 μ M) were tested at speeds between 10,000 and 16,000 rpm for a total run time of 20 h. Data were collected as described above.

Analyses of the SE data were performed with Ultrascan II (Demeler, 2005) to obtain an averaged molecular mass for the data acquired under each condition. The data were globally fitted to a non-linear regression model for a single species, non-interacting system, using the following equation:

$$\ln Cr = Mw \omega^2 \frac{(1 - \bar{v}\rho)}{2RT} (r_2^2 - r_1^2) + \ln Cm$$

where Cr is concentration change, Mw is molecular mass, ω^2 is angular velocity, \bar{v} is the partial specific volume, ρ is the density of the solvent, R is the gas constant, T is the absolute temperature, r_1 and r_2 are the radial distances from the rotor axis to the meniscus and the bottom of the cell, respectively and Cm is the concentration at the meniscus.

SEDNTERP v1.05 (Laue *et al.*, 1992) was used to determine buffer density (1.002 g mL⁻¹) in addition to \bar{v} for AtIPK1 (0.739 mL g⁻¹). These parameters were included in the final model to solve the calculated molecular mass. Theoretical masses of

AtIPK1 (~50 kDa) and dsDNA (~6 kDa) were calculated to aid comparison with ProtParam (ExpASy) and OligoCalc (Kibbe, 2007) respectively.

2.2.3 Fluorescence lifetime imaging microscopy Förster resonance energy transfer

Fluorescent lifetime imaging microscopy Förster resonance energy transfer (FLIM-FRET) in *N. benthamiana* was performed following a recently published protocol (Camborde *et al.*, 2017). Here the methods are reported with some modifications. Briefly, the method involves loading of donor probe-expressing (typically protein-eGFP family) tissue with an acceptor probe and imaging of the interaction thereof.

2.2.3.1 Agroinfiltration of *N. benthamiana* leaves

N. benthamiana seeds provided by Hayley Whitfield were grown in short day light cycle conditions (8 h light at 21 °C and 16 h dark at 18 °C) for 4 weeks prior to infiltration. For each construct ≥ 3 leaves of ≥ 2 plants were infiltrated.

Stocks of *Agrobacterium tumefaciens* GV3101 PMP90 strains containing *pEAQ AtIPK1-GFP* and *pEAQ GFP* were also kindly provided by Hayley Whitfield. The *pEAQ* vectors were designed to allow high levels of transient expression of proteins in plants (Sainsbury *et al.*, 2009). *A. tumefaciens* strains were inoculated into 10 mL Lennox-LB (10 mg/mL tryptone, 5 mg/mL sodium chloride, 5 mg/mL yeast) containing gentamicin (25 μ g/mL) and kanamycin (50 μ g/mL) for selection, and grown for 48 h at 30 °C, with 200 rpm shaking. The cultures were pelleted at 2000x g, 15 °C for 15 min and the supernatant was discarded. Pellets were resuspended in 10 mL infiltration solution [150 μ M 3:5;-dmethoxy-4'-hydroxyacetophenone, 2% sucrose in 1x Murashige and Skoog media (Murashige and Skoog, 1962)]. Suspensions were left at rt for 5 h before *N. benthamiana* leaves were infiltrated on the abaxial side using a 1 mL needleless syringe. Plants were returned to the growth rooms for 1-5 days (in short day light cycle conditions described above) after which they were screened for GFP expression every 24 h using an UV box.

2.2.3.2 Sample fixation, permeabilization and RNase treatment

Two days after GFP infiltration, leaf discs were harvested with 6 mm cork borer. Fixation, permeabilization and staining protocols were followed according to Camborde *et al.* (2017). In alternative approaches to protocol sample fixation, a vacuum desiccator was used with leaf discs submerged in solution within a 24 deep well plate. A vacuum was applied for 2 to 3 min and released slowly 10 to 20 times, until leaf discs became transparent. Whether using a needleless syringe or vacuum desiccator, an alternative fixative solution was trialled by addition of 0.1% triton X-100. When using the triton X-100 there was no need for the succeeding step of leaf permeabilization with proteinase K.

2.2.3.3 SYTOX Orange staining

In order to stain the leaf nuclei with the FRET acceptor Sytox Orange (SO), vacuum infiltration was performed with 2-3 discs using a 5 mL syringe with 2 mL ice cold SO solution (5 μ M SO in 1x TBS), the vacuum was applied and released slowly 10 to 20 times. Alternatively, leaf discs submerged in 5 mL SO solution within a deep well plate placed into a vacuum desiccator, where a vacuum was applied 10-20 times for 2-3 minutes with an electric pump and slowly released. The leaf discs were finally washed a further three times and kept in 1x TBS for use in microscopy.

2.2.3.4 Fluorescence microscopy

For initial verification of reporter expression, leaf discs were imaged on Zeiss Axioplan 2ie upright epifluorescence microscope fitted with a colour Zeiss AxioCam camera. Samples were visualized using a 10x objective and captured using Zeiss Axiovision software at rt. Illumination was provided by a mercury lamp (X-Cite 120Q, Lumen Dynamics) and SO (excitation 546/12, emission 590) or GFP (excitation 470/20, emission 505-530) filter sets. Final images were processed using MS PowerPoint.

2.2.3.5 Fluorescence lifetime imaging microscopy

Two-photon microscopy was performed in the Henry Wellcome laboratory for Cell imaging, UEA, using a LaVision Biotec TriMScope II mounted on a Zeiss Observer D.1. Images were collected with a 20x (0.8 NA) PlanApoChromat objective lens and ImSpector Pro software (LaVision Biotec). GFP and SO were excited at 930 nm (Coherent Vision II Ti:Sapphire laser) and fluorescence emission collected using two, non-descanned GaAsP detectors at 525 ± 25 nm (GFP) and 565 ± 12 nm (SO). For lifetime imaging (FLIM), the GFP fluorescence was directed to a 16-channel TCSPC FLIM detector using a high-NA liquid light-guide. Lifetime was measured over an intra-pulse period of 11.6 ns with 80 fs time bins. FLIM images typically required averaging 10 to 20 images with a total capture time of around 10 s.

2.2.3.6 FLIM-FRET analysis

Downstream processing of FLIM images were performed in ImSpector Pro (LaVision Biotec). For each nucleus within a single image, a measurement of average lifetime of GFP was obtained by drawing an object of interest and using a mono or bi-exponential model to find the best fit. The mean τ of GFP in the presence or absence of SO was then used to calculate FRET efficiency using the following equation:

$$FRET \text{ efficiency (\%)} = 1 - \left(\frac{\text{average } \tau \text{ donor} + \text{acceptor}}{\text{average } \tau \text{ donor}} \right)$$

The data were not suitable for further statistical analyses.

2.3 Results

2.3.1 Purification of AtIPK1

To conduct *in vitro* DNA binding assays, the production of high purity AtIPK1 is an essential starting point. This was achieved by large scale heterologous expression in *Escherichia coli* Rosetta™(DE3) cells. *E. coli* is a convenient expression system to work with due to speed of growth, inexpensive growth media and ease of genetic manipulation (Chen, 2011). There are many commercially available strains which have specific genotypes to be used for specialised purposes, Rosetta™(DE3) host strain has been designed to enhance eukaryotic protein expression, and supply tRNA for codons rarely used in *E. coli*.

The full length *AtIPK1* gene had previously been cloned (Whitfield, 2013) into the pOPINF vector (addgene plasmid #26042), one of a suite of pTriEx2-derived vectors designed for compatibility with In-fusion™ cloning (Berrow *et al.*, 2007). Particular features of this plasmid are the *T7lacO* promoter for inducible gene expression, restriction sites (*KpnI* and *HindIII*) for aid of cloning, ampicillin resistance for selection, an N-terminal 6x His affinity tag for purification with a 3C protease site for tag removal and compatibility with *E. coli* and eukaryotic hosts. Accordingly, the protein purification steps involved nickel affinity chromatography (Appendix 2), followed by 3C cleavage and removal of the 6xHis tag (Figure 2.6). The last step of purification was achieved using size exclusion chromatography to improve homogeneity of protein. In a typical 6 L production, we were able to yield approx. 3 mL of 12 mg/mL AtIPK1.

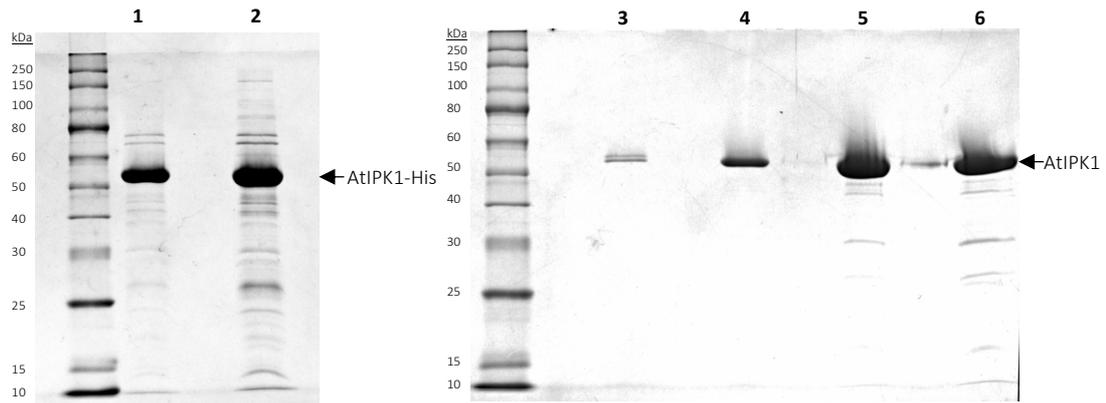


Figure 2.6 Example of SDS-PAGE analysis of protein samples from AtIPK1 purification. Fractions collected from different purification steps, alongside protein ladder; arrows show purified AtIPK1-His/AtIPK1 (~50 kDa). Lanes: 1, AtIPK1-His protein from Ni-NTA purification; 2, AtIPK1-His concentrated protein; 3, AtIPK1-His protein + 3C reaction mixture; 4, His-cleaved AtIPK1 purified by Ni-NTA; 5, AtIPK1 concentrated from NiNTA column; 6, AtIPK1 from size exclusion purification.

2.3.2 AtIPK1 binds to DNA in FP assays

In direct follow-up to the previous work of a colleague investigating AtIPK1-DNA interaction with FP (Whitfield, 2013), binding assays were repeated using the same binding reaction conditions and oligomer target dsDNA sequence (referred to as GC), for proof-of-concept. AtITPK4, a cytosolic protein, of the inositol tris/tetrakisphosphate class (Sweetman *et al.*, 2007) was used as a negative IPK control in tandem. An increase of probe polarization was observed as higher amounts of AtIPK1 were titrated into the binding reaction (Figure 2.7), indicative of protein-DNA binding and in clear contrast to AtITPK4, in which polarization values stayed constant. The wide dynamic range of the assay and tight fitting is similar to the binding of inositol phosphate probe 2-FAM-InsP₅ to AtIPK1 (Whitfield *et al.*, 2018) and the EC₅₀ value of 63 nM is comparable to that obtained with 2-FAM-InsP₅, ~529 nM (Chapter 3.3.4). While the structure of AtITPK4 has only recently been solved (Whitfield, unpublished), the protein does not obviously possess zinc-co-ordinating residues that are the hallmark of some DNA-binding domains. Utilizing the easy set-up of FP, further assays were designed to inspect preferred DNA binding targets and non-specific DNA binding.

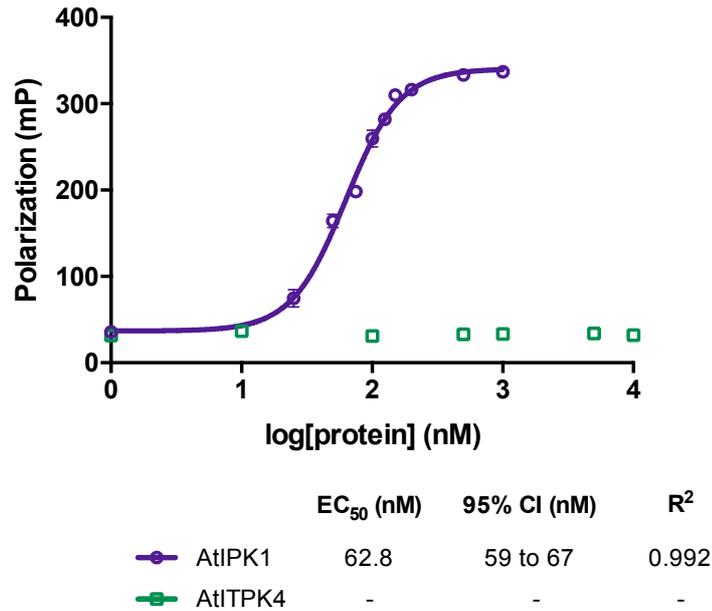


Figure 2.7 AtIPK1 (purple, open circles) binds to FAM-labelled DNA in FP assay. Variable slope curves generated by Prism v6.0 (Graphpad) showing the mean \pm SD of fluorescence polarization data from AtIPK1 interacting with 2 nM of GC (5'-FAM-CCCCGGGTACCGAGCTCGAATTCAGTGGCCGTCGTTTTAC-3'). These binding assays were performed in 20 mM HEPES pH 7.5, 1 mM MgCl₂, and incubated for 1 h before plate readings at 25 °C. AtITPK4 (green, open squares) is included as a negative control.

Ionic components of binding buffers can affect the binding parameters observed in protein-DNA interactions (Lohman and von Hippel, 1986). Salt is often varied in such experiments to limit electrostatic predominant points of contact, the non-specific interactions (Hart *et al.*, 1999). Preliminary FP assays were conducted with 0-100 mM NaCl to determine the optimum concentration of salt to use in later assays. The general influence of NaCl additions was observable as a right shift of standard curves (Figure 2.8) and a dramatic increase of EC₅₀ values, due to NaCl preventing non-specific interactions between AtIPK1 and FAM-oligomer. Upon the addition of 50 mM NaCl, the EC₅₀ increased almost five-fold (from ~62 nM up to 281 nM). At the highest salt concentration (100 mM NaCl) the standard curve did not reach saturation, and therefore further analyses were not completed. Transformations of the raw polarization data from FP experiments were used to calculate the fraction of bound protein-DNA complex for a kinetic analysis. Prism v6.0 (Graphpad) built in one site – total binding

model was used to obtain the dissociation constant K_d for AtIPK1-DNA interaction under these specified conditions (Figures 2.8b and 2.10b). Moving forwards, 50 mM NaCl was used in the buffer for *in vitro* binding experiments.

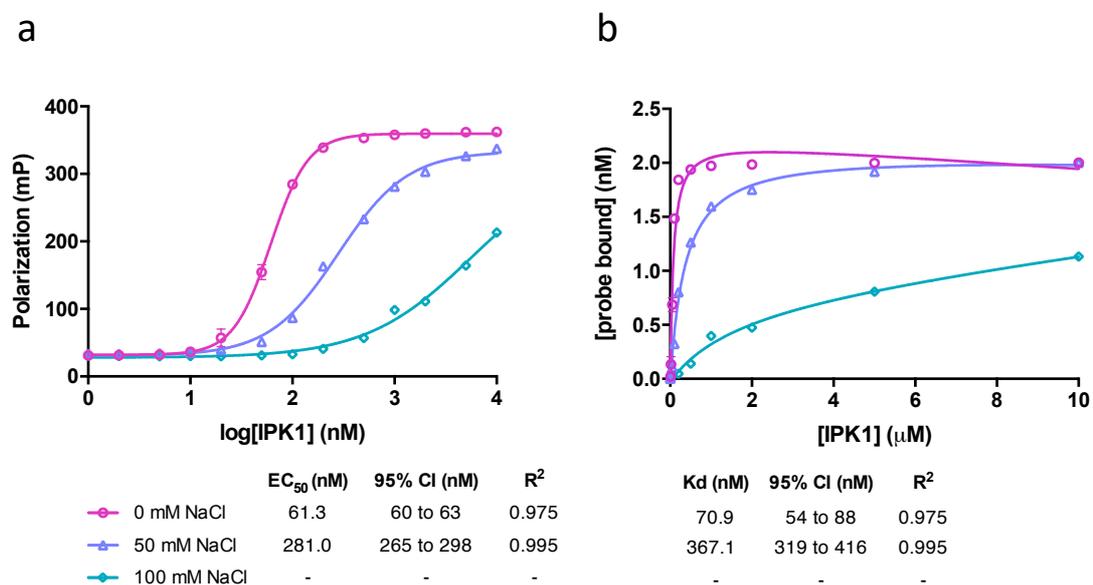


Figure 2.8 Salt dependence of AtIPK1-DNA interaction monitored by FP with 2 nM FAM-AT 40mer. Each point is the mean \pm SD shown as error bars under different salt conditions: 0 mM NaCl as pink open circles, 50 mM NaCl as purple triangles and 100 mM NaCl as teal open diamonds. FP data was processed with Prism v6.0 (GraphPad) using (a) a variable slope model and (b) a one site – total binding model using fraction of bound AtIPK1 with different oligomers to allow calculation of K_d .

DNA binding proteins are commonly characterized in respect of their interaction with known or putative target binding sequences, and commonly by EMSA or F-EMSA. To further validate the novel *in vitro* FP-approach shown in Figures 2.7-2.8, and in the absence in the literature of known DNA-binding or target sequences for AtIPK1, three pairs of FAM-labelled oligomers were annealed to act as putative binding partners of AtIPK1 for fluorescence-based assays. These were: FAM-labelled GC rich 40-nt oligomer (FAM-GC 40mer), FAM-labelled AT rich 40-nt oligomer (FAM-AT 40mer) and FAM-labelled 12-nt oligomer (FAM-12mer) [see Chapter 3]. The oligomers were run on a 20% NATIVE gel to verify annealing, or otherwise, using ethidium bromide to detect the DNA. The gel visualised under UV light revealed that the FAM-GC and AT 40mers were

annealed (Figure 2.9), whereas the FAM-12mer bands remained ambiguous. The brightness of the bands varies per lane, the double stranded DNA which travels a shorter distance through the gel matrix appears crisp and bright, however single stranded DNA and DNA that migrated further into the gel are blurry and harder to distinguish. The AT-rich palindromic 12mer sequence is short and prone to form primer dimers, which may account for the apparent failure of annealed DNA or low annealing efficiency. This evidence shows at least two populations of dodecamer, perhaps most likely FAM-labelled and unlabelled ss-probe. Future annealing attempts were made using only the labelled dodecamer, in order to limit the potential populations to ss and ds FAM-12mer.

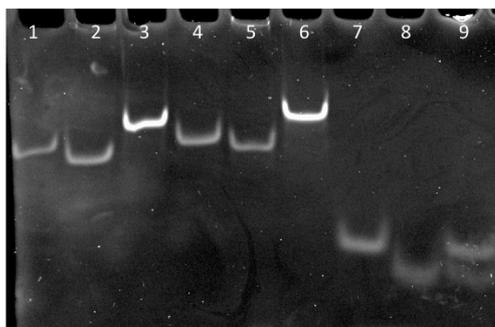


Figure 2.9 Annealed nucleotide oligomers run on 20% polyacrylamide gel stained with ethidium bromide, showing single stranded vs double stranded DNA used in DNA binding assays. Lanes: 1, ssFAM-GC; 2, ssGC R; 3, annealed FAM-GC; 4, ssFAM-AT F; 5, ssAT R; 6, annealed FAM-AT; 7, ssFAM-12-mer; 8, ss 12; 9, annealed FAM-12mer.

FP experiments with annealed FAM-oligomers were used to detect whether AtIPKI had greater affinity between AT or GC rich DNA. The curves for FAM-GC and AT 40mers are almost identical (Figure 2.10), with respective EC₅₀ values of 311.4 nM (95% CI, 292.6 to 331.5 nM) and 281.2 nM (95% CI, 265.2 to 298.1 nM). This suggests a possible 9% higher affinity between AtIPKI and AT rich oligomer compared to GC, however limitations in accuracy of this method mean it is likely there is no significant difference. The corresponding *K_d* estimates (415.4 ± 28.8 nM and 367.1 ± 24.2 nM) reflect this. In the case of the 12-mer, the curve differs slightly, with a lower *K_d* 337 nM (EC₅₀ 235.8 nM; 95% CI, 213 to 262 nM) and smaller dynamic range.

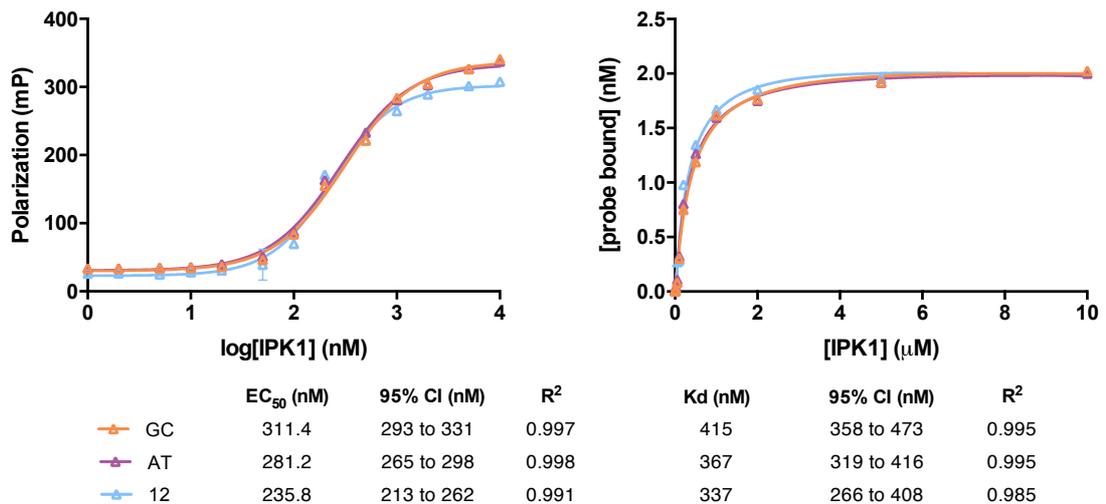


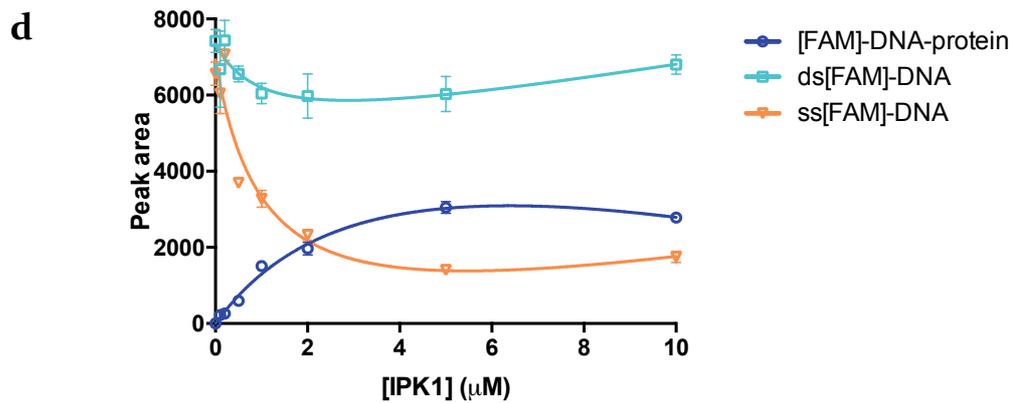
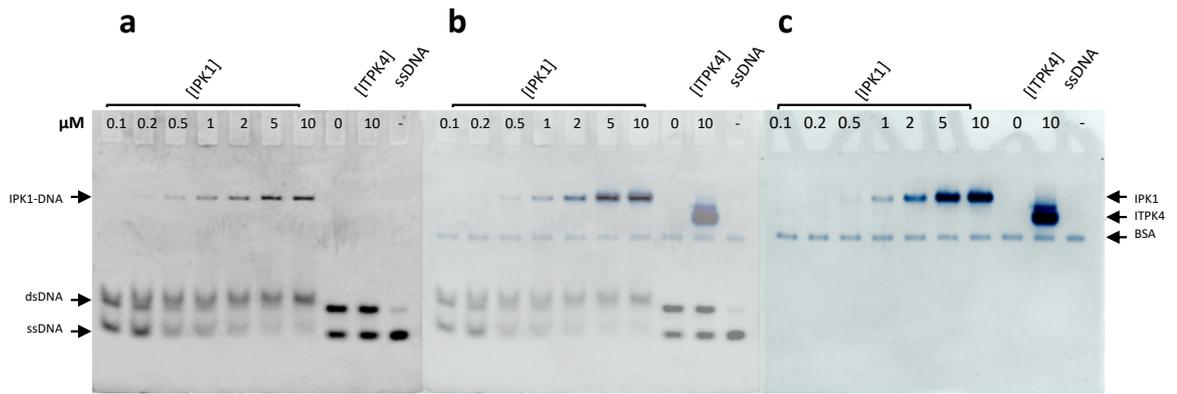
Figure 2.10 AtIPK1 binds to three different FAM-oligomers. Data shown are the mean \pm SD of FP data from AtIPK1 interaction with 2 nM of FAM-GC 40mer (orange, open triangles), FAM-AT 40mer (pink, open triangles) and FAM-12mer (blue, open triangles) probes. Curves were generated in Prism v6.0 (Graphpad). Raw FP data was processed using (a) a variable slope model of AtIPK1 binding FAM-oligos and (b) a one site – total binding model using fraction of bound AtIPK1 with different oligomers to allow calculation of K_d . These binding assays were performed in 20 mM HEPES pH 7.5, 1 mM MgCl₂ and 50 mM NaCl, and incubated for 1 h before plate readings at 25 °C.

2.3.3 F-EMSA

Fluorescence-based electrophoretic mobility shift assays (F-EMSA) are a widely used assay of DNA-binding and were utilized here to support the evidence of AtIPK1-DNA interaction detected from FP assays described above. In this study, AtIPK1 can be seen to interact with the FAM-labelled probe on the 8% native gel (Figure 2.11a-c), as the characteristic shift of the free DNA co-aligns with the Coomassie stained AtIPK1. The intensities of the DNA-bound and free bands reflect the equilibrium of the complex formation and quantitative analysis was attempted using the densitometry function in ImageJ (Figure 2.11d-e).

Using the annealed FAM-12mer brought complexity to the F-EMSA analysis. Firstly, the free FAM-12mer negative control is clearly two separate bands, indicative of two populations of DNA. These were predicted to be single stranded and annealed FAM-12mer, of which ssDNA was identified by running an

additional sample of ss FAM-12mer, seen in Figure 2.11a. This is problematic because the free probe bands are crisp and clear in the absence of AtIPK1, but both appear to shift and blur upon the addition of 100 nM protein. As the protein titration increases, the band corresponding to ssDNA reduces preferentially, over the ds probe. This perhaps suggests the preference of ssDNA binding of AtIPK1. Quantification in this case is more challenging, however using densitometry analysis, the dsDNA band was shown to remain approximately constant (Figure 2.11d). This band was therefore used as a point of reference to the AtIPK1-bound DNA and ssDNA to quantify the relative densities per lane (Figure 2.11e). This analysis shows that ssDNA reduces to approximately 23% from 88% (of total signal in the lane) in the presence of 5 μ M AtIPK1, however the bound DNA band increases from 0 to 51% relative intensity. This adds up to unaccounted for fluorescence, but perhaps more surprising is the small decrease of relative intensity with 10 μ M AtIPK1 (41%). A further complication to the analysis is afforded if one considers the possibility that AtIPK1 interaction with DNA alters the equilibrium of ss to ds DNA. Undoubtedly, the ss pool is depopulated by protein addition across the entire titration, while ds pool depopulation is most apparent at lower protein concentration. The small decrease of protein-bound probe at 10 μ M AtIPK1 may point to inaccuracies of the method of measurement, loading the gel or possibly the dissociation of the AtIPK1-DNA complex. Nevertheless, it is clear, again, that AtIPK1 binds DNA.



[IPK1] (μM)	0	0.1	0.2	0.5	1	2	5	10
% bound DNA*	0.0	3.3	3.6	9.1	25.0	33.0	50.6	40.9
% free ssDNA*	88.4	90.4	95.1	56.4	54.2	39.0	23.3	25.7

* relative to dsDNA detected per lane

Figure 2.11 F-EMSA gel analysis of AtIPK1 interaction with DNA. (a-c) images of an 8% native gel with 5 nM FAM-DNA with increasing amounts of purified AtIPK1 in binding reaction (0.1 mg/mL BSA, 1 mM EDTA, 20 mM HEPES pH 7.5, 1 mM MgCl₂, 50 mM NaCl), plus negative controls of 10 μM ITPK4 and non-annealed ssFAM-DNA, visualised by (a) FAM detection with Typhoon FLA9.1000, (b) image editing and overlaying with 0.6% opacity in GIMP and (c) protein staining with InstantBlue™; (d) plot of non-linear regression analysis of densitometry data of F-EMSA bands from ImageJ and; (e) relative % of protein-bound DNA and free ssDNA compared to the amount of dsDNA in each gel lane. The colour gradient from red (low %) to green (high %) shows the abundance of FAM probe detected in each lane.

2.3.4 AtIPK1 is a dimer in solution

To investigate the native and DNA-bound form of AtIPK1, analytical ultracentrifuge sedimentation equilibrium (AUC-SE) experiments were used. This method requires purity of component species and monitoring of samples over the course of several tens of hours. Therefore, the ideal buffer should be conducive to the maintenance of protein or DNA stability in solution, to prevent non-ideal systems obstructing the analyses.

Initially, the oligomeric state of native AtIPK1 was assessed by SE in the buffer conditions used in earlier assays (20 mM HEPES pH 7.5 and 1 mM MgCl₂) with the addition of 50 mM NaCl to reduce non-specific DNA-protein interactions. This data is summarised in Table 2.3, where the calculated average molecular mass shifted from ~ 85 kDa to 200 kDa between 20 and 40 h, suggesting protein aggregation. In tandem, degradation of AtIPK1 was observed over 48 h at rt (Figure 2.12). Contaminant profiles are not dramatically different over time monitored, however the AtIPK1 band appears reduced at 48 h and more contaminant bands are present between ~37 and ~25 kDa.

Table 2-3 The calculated average molecular mass of AtIPK1 from SE experiments.

Protein in 20 mM HEPES pH 7.5, 1 mM MgCl₂ and 50 mM NaCl monitoring A280 at 20 °C, the theoretical molecular mass of monomeric AtIPK1 is ~50 kDa.

IPK1 concentration (μM)	Average molecular mass (kDa) at 20 h	Average molecular mass (kDa) at 40 h
5	80	[*]
10	86	198
15	88	218

*data not suitable for analysis

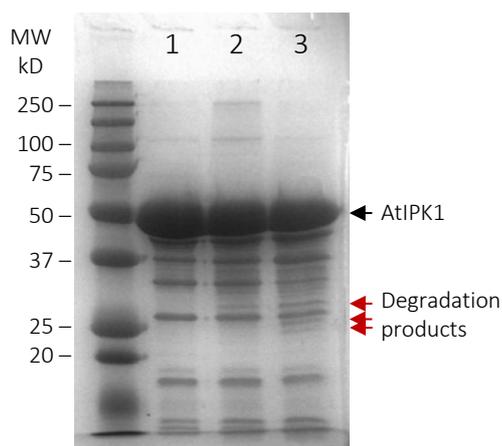


Figure 2.12 SDS-PAGE analysis of AtIPK1 stability run in parallel to AUC-SE analysis. AtIPK1 prep is overloaded onto 12% SDS-PAGE gel, at timepoints: 1 = 0 h, 2 = 24 h, 3 = 48 h. AtIPK1 is indicated by black arrow; protein bands that are considered degradation products are indicated by red arrows.

To improve AtIPK1 stability over the course of the SE analysis, a trial with five alternative buffers (Table 2.4) was conducted. A single-species, non-interacting model was used to fit the data and calculate the average molecular mass. The raw SE curves (for example see Figure 2.13a) could be fitted to a single species model more accurately in conditions 2, 4 and 5. The addition of reductants DTT and TCEP (in buffers 4 and 5 respectively) improved the stability of the potential AtIPK1 homodimer (~100 kDa) after 20 h. This route was preferred over the use of Tris (buffer 2), which also provided relatively stable conditions, to more closely reflect the environment used in other binding experiments. For an ideal single-species system, a plot of the natural log of A280 against the square of radial position is linear, with the slope proportional to the molecular weight of the molecule. In the transformations of this SE data, improved linearity can be seen with the addition of DTT (Figure 2.13b), however both conditions are a poor fit when considering the residual plots. Consequently, DTT was added to the binding buffer in subsequent SE experiments at a lower concentration (0.5 mM) than in buffer 4, to minimise interference from oxidized DTT at A280. Furthermore, SE analysis using improved buffer conditions similarly suggest that AtIPK1 is a homodimer (~93 kDa) in solution (Table 2.5).

Table 2-4 SE analysis of 10 μ M AtIPK1 in different buffering solutions. Protein samples, with theoretical monomeric mass of \sim 50 kDa (ProtParam), were centrifuged at 10,000 rpm rotation speed, monitored at A280 over 40 h, at 20 $^{\circ}$ C.

Buffer no.	Buffer composition	Average molecular mass (kDa)
1	20 mM HEPES pH 7.5, 50 mM NaCl	160
2	20 mM Tris pH 8.0, 50 mM NaCl	116
3	20 mM HEPES pH 7.5, 50 mM NaCl, 1 mM MgCl ₂	220
4	20 mM HEPES pH 7.5, 50 mM NaCl, 1 mM MgCl ₂ , 2 mM DTT	115
5	20 mM HEPES pH 7.5, 50 mM NaCl, 1 mM MgCl ₂ , 1 mM TCEP	86

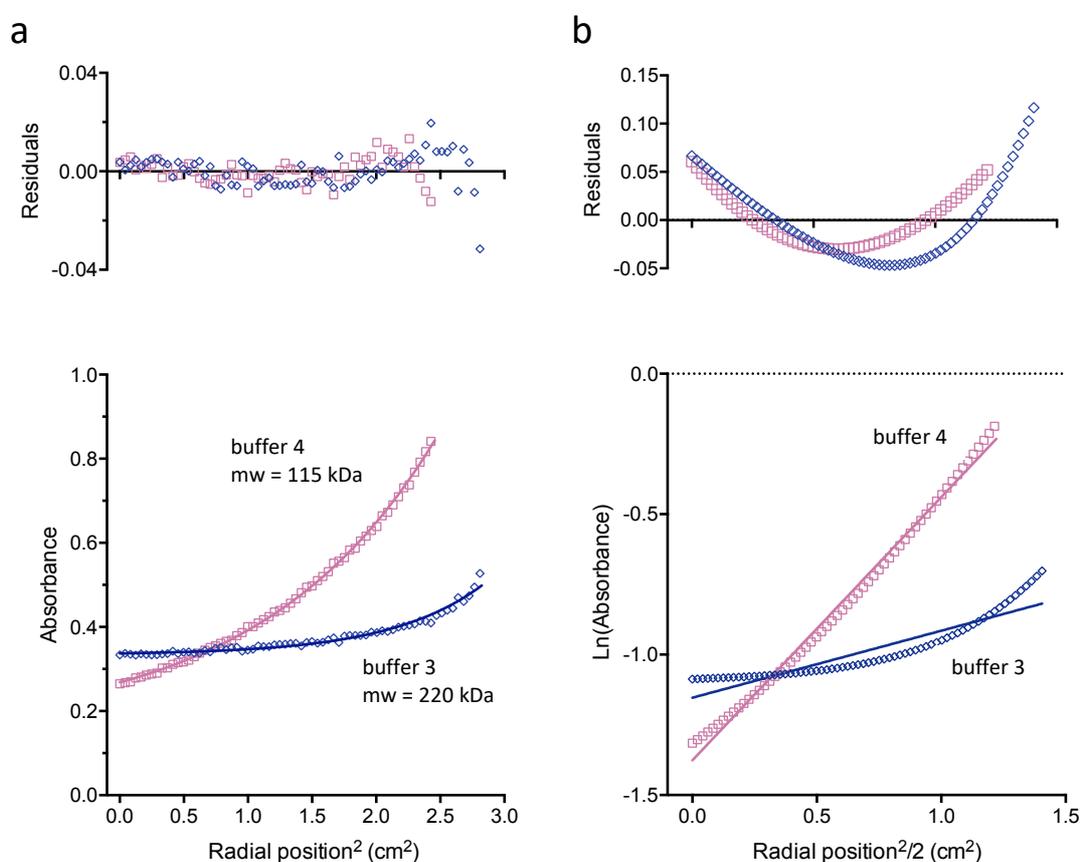


Figure 2.13 Two representative SE curves (out of five) for AtIPK1 with (open squares, pink) and without (open diamonds, blue) addition of 2 mM DTT in 20 mM HEPES pH 7.5, 1 mM MgCl₂, 50 mM NaCl buffer. The observed data at A280 are shown with the final fitted curve superimposed. (a) The absorbance profiles of 10 μ M AtIPK1 at A280 as a function of radial position and (b) the data replotted as the natural logarithm of A280 against the square of the radial position, and linear regression analysis. Residuals between fitted curves and the experimental data are displayed in the upper panels.

Table 2-5 Calculated molecular mass determined from SE with AtIPK1, with theoretical monomeric mass of ~50 kDa. SE was performed in 20 mM HEPES pH 7.5, 1 mM MgCl₂, 50 mM NaCl buffer and 0.5 mM DTT at 11,000 rpm, monitoring A280, for 25 h at 20 °C.

IPK1 concentration (μM)	Average molecular mass (kDa)
8	100
12	86
16	89

SE was finally used to investigate the AtIPK1-DNA complex by following the change of molecular mass at A260, to monitor DNA absorbance rather than protein (Table 2.6). A single-species, non-interacting model was used to calculate for DNA alone, yielding ~8 kDa, this molecular mass is slightly higher than the ~6 kDa oligonucleotide estimated by OligoCalc (Kibbe, 2007) from the nucleotide sequence. However, since the \bar{v} used to calculate the molecular mass change for DNA was taken from the protein, the calculated mass is not expected to be accurate. The micromolar ratios of protein and DNA were varied to test whether changes in molecular mass calculations could be observed, to provide information about the stoichiometry of the interaction. Upon the addition of AtIPK1, the calculated molecular mass dramatically shifted up to ~60 kDa, suggesting that the protein is binding DNA in a monomeric form. In this analysis, the single-species, non-interacting model was a bad fit for samples containing higher concentrations of protein, this may be due to mixed populations of bound and non-bound protein-DNA complex or from errors introduced by peak interference from the protein absorbance.

Table 2-6 Calculated molecular mass determined from SE analysis of AtIPK1 and DNA in different micromolecular ratios. The theoretical monomeric mass of AtIPK1, ~50 kDa (ProtParam) and dsDNA, 6 kDa (OligoCalc).SE was performed in 20 mM HEPES pH 7.5, 50 mM NaCl, 1 mM MgCl₂, 0.5 mM DTT, at 10,000 following A260 for 24 h at 20 °C.

DNA-protein ratio (μM)	Average molecular mass (kDa)
2:0	8
2:4	62*
2:6	63*
1:2	46
1:4	75*

* poor fit of data seen through residual plots

To diminish potential peak interference at A260 from AtIPK1, SE was performed with lower protein concentrations (Table 2.7). Although, once again these are not accurate calculations of the AtIPK1-DNA complex due to the differences of \bar{v} between the two molecular species. Nevertheless, a large increase in molecular mass of the monitored DNA in the presence of protein indicating the formation of a protein-DNA complex. Whether in mixed solution with excess DNA, excess protein or with 1:1 ratio of DNA and AtIPK1 counterparts, the average calculated molecular masses of the heterodimer complex remain similar (within a range of between 43 to 54 kDa) when the protein concentration does not exceed 2 μM . These calculated masses indicate that the monomeric form of AtIPK1 binds to DNA, as we may expect a protein-DNA complex containing dimeric AtIPK1 to be much larger i.e., twice the size.

Table 2-7 Calculated molecular mass determined from SE analysis of AtIPK1 and DNA in different micromolecular ratios. where AtIPK1 has a theoretical monomeric mass of 50 kDa and dsDNA 6 kDa. SE experiments performed in 20 mM HEPES pH 7.5, 50 mM NaCl, 1 mM MgCl₂, 0.5 mM DTT, at 16,000 following A260 for 20 h at 20 °C.

DNA-protein ratio (μM)	Average molecular mass (kDa)
2:0	12.5
2:2	52
4:2	46
3:2	54
1:1	43

2.3.5 Preliminary data from FLIM-FRET

The foregoing experiments provide evidence of binding of DNA to AtIPK1 that could explain the nuclear localization of AtIPK1 observed with GFP-tagged protein (Whitfield, 2013; Kuo *et al.*, 2014; Adepoju *et al.*, 2019). In order to confirm that nuclear localisation involves DNA binding, a series of FLIM-FRET experiments were performed. The advantage of this approach over confocal microscopy observed co-localization arises from the principle that energy transfer between molecular orbitals of donor and acceptor is inversely proportional to the sixth power of the distance between the two (Müller *et al.*, 2013), highly indicative of interacting molecules. Towards this goal, agroinfiltration of *Nicotiana benthamiana* plant leaves with pEAQ-HT vectors containing *AtIPK1* with C-terminal *GFP* tag and *GFP* alone as a standard for comparison were used to conduct FLIM-FRET with Sytox Orange as an acceptor, following a recently published protocol (Camborde *et al.*, 2017). Fortuitously, these particular constructs had previously been cloned and transformed into *Agrobacterium* and were readily available to use with thanks to Hayley Whitfield.

The expression of *GFP* and *AtIPK1-GFP* were primarily detected in discreet patches of leaf tissue of *Agrobacterium*-infiltrated plants by visualization using an UV box. The infiltration technique floods the apoplast, the intercellular spaces, with *Agrobacterium* and gives preferential access to the epidermal layer and the mesophyll tissue immediately sub-epidermal. Plants were monitored

over five days. Although ubiquitous expression within the leaf tissue was not achieved, higher levels of GFP fluorescence were detected after 48 h, therefore subsequent protocol steps were performed after a two-day waiting period. Once GFP fluorescence was confirmed, leaf discs were harvested and prepared so that tissues treated with or without Sytox Orange (SO) could be compared by FLIM.

Fluorescence of SO, the nucleic acid dye (Yan *et al.*, 2002), was mostly observed peripherally in the leaf disc and commonly in the nuclei of guard cells present in the leaf epidermis. This distribution severely limited the overlap of donor and acceptor required for FRET analysis (Figure 2.14). Improvement of SO infiltration was attempted by using different fixatives and by performing the experiments using an alternative infiltration method with a vacuum desiccator. These attempts were of limited success where treatments with 0.1% triton additions slightly improved the depth of perimeter SO staining. Despite this, a small population of nuclei of sub-epidermal and mesophyll cells showing co-localisation of FRET donor and acceptor were observed and FLIM measurements were taken.

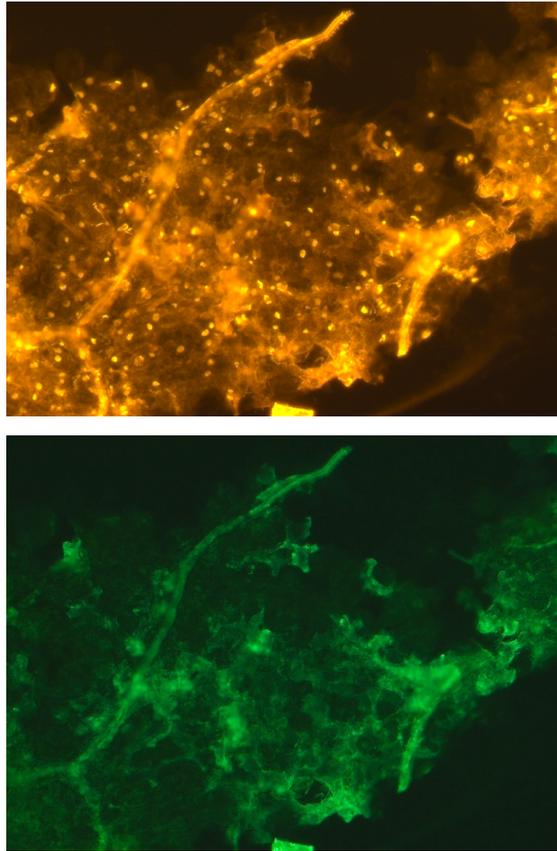
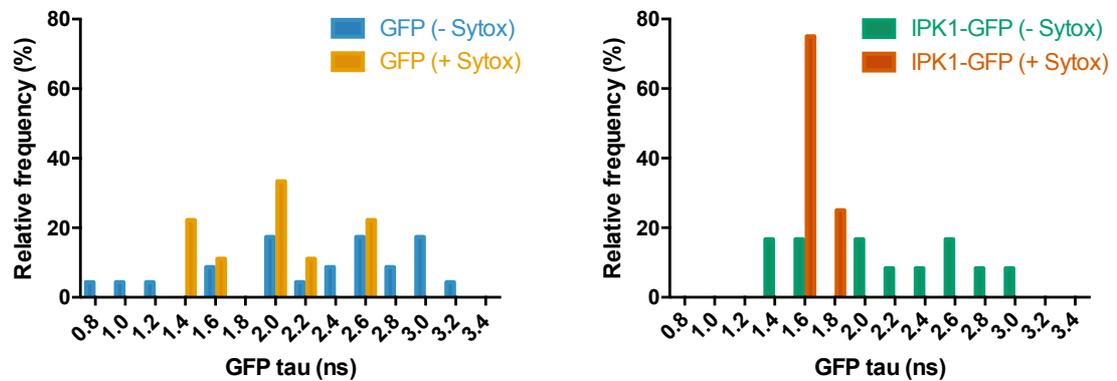


Figure 2.14 Microscopy images of GFP-transformed *N. benthamiana* leaf discs stained with SO visualized by two filters: SO (top) and GFP (bottom). Despite infiltration of SO into nuclei of peripheral leaf tissues, the dye did not enter the entire samples. When examining SO stained nuclei for overlap with GFP fluorescence, there was no co-localization.

FLIM measurements of GFP were acquired using the ImSpector Pro software (LaVision Biotec). Within the samples of non-treated leaf discs, a large range of GFP lifetimes were observed (Figure 2.15), causing large margins of error. In contrast, the SO treated leaf discs had a lower degree of variation of GFP lifetime, especially the AtIPKI-GFP construct. The lifetime (τ) data was best fit using a bi-exponential model and the mean τ was averaged among repeat measurements for each construct. According to the Camborde *et al.* (2017) protocol, the average donor τ was used to calculate FRET efficiency for AtIPKI-GFP-SO and GFP-SO pairs (~25.5% and ~13.3% respectively). Although non-parametric tests can bypass some concerns of non-normally distributed data and differences in sample sizes; they cannot account for variance nor different shapes of sample

distribution. Therefore, it was not appropriate to use statistical analyses for this preliminary data to generate *P* values to measure differences between the means.



Donor	Acceptor	τ	SEM	N	E
GFP	-	2.28	± 0.14	23	
	+	1.98	± 0.16	9	13.3%
IPK1-GFP	-	2.15	± 0.17	12	
	+	1.60	± 0.03	8	25.5%

Figure 2.15 Distributions of average lifetime (τ) measurements of GFP and AtIPK1-GFP transiently expressed in *N. benthamiana* leaf nuclei. These data were collected over the course of three months, and includes data from samples prepared with 4% paraformaldehyde, 50 mM sodium cacodylate and with additional 0.1% triton in the infiltration solution.

2.4 Discussion

In agreement with findings from preliminary FP experiments in the Brearley lab, DNA-binding activity of AtIPK1 was observed using *in vitro* methods: FP, F-EMSA and AUC. In this study, we show that AtIPK1-DNA binding is salt dependent and non-sequence specific through FP. The stability of the AtIPK1-DNA complex was demonstrated using F-EMSA, revealing that AtIPK1 could not fully sequester the non-specific probe and suggesting a preference for ssDNA. The oligomeric state of AtIPK1 in solution was shown to be dimeric through AUC-SE, however when monitoring AtIPK1 with unlabelled DNA, the molecular weight of the AtIPK1-DNA complex suggests that AtIPK1 interacts with DNA in either monomeric or dimeric form. We could not unequivocally confirm the AtIPK1-DNA binding activity using FLIM-FRET *in planta*, however preliminary data does not discount the potential physiological relevance for this novel function of AtIPK1.

2.4.1 AtIPK1 binds DNA *in vitro*

During this investigation, equilibrium binding constants and dosage-dependence values were obtained to aid characterization of AtIPK1-DNA binding using FP. Within this study, these were used to compare DNA binding affinity in different conditions, such as salt additions and oligomer preferences. Whilst F-EMSA with AtIPK1 and DNA, requiring less protein, supported the *in vitro* interaction visually, it only provides a snapshot of binding equilibrium within the gel matrix; a more complex system. As a qualitative assay, subtle changes in DNA binding were difficult to validate. The protein-DNA complex can disassociate and re-associate within the matrix and less stable complexes, which may arise through additional interactions and over extended periods of electrophoresis, can be seen through smeared bands (Figure 2.11). Alternately, protein-DNA complexes can be artificially stabilized within the matrix relative to free solution, and this can be assisted with additives such as EDTA.

At the time of writing, there are no known IPK DNA binding proteins to allow direct comparison of AtIPK1 binding activity. In this study, the non-nuclear AtITPK4 was used alongside AtIPK1 in FP and F-EMSA experiments as a negative

control for DNA binding. Despite low sequence similarity (blastp: no similarity detected) between AtIPK1 and AtITPK4, they provide a point of comparison as similar sized IPKs (~50 kDa and ~55 kDa respectively) which both have a positively charged InsP binding pocket. The InsP binding pocket of AtITPK4 however was not capable of binding DNA non-specifically, in the absence of salt. This suggests that the binding of DNA to AtIPK1 is not necessarily a result of the presence of an InsP binding pocket.

In the wider context of known plant DNA binding proteins, equilibrium constants can be used to draw comparisons for reference. For example, we obtained K_d values of 10^{-7} M affinity for non-specific binding of AtIPK1 under manipulated conditions, which fall within reports for DNA binding affinity of plant transcription factors with target oligomer sequences have been reported K_d values of 10^{-12} to 10^{-6} M (Hao *et al.*, 1998; Welner *et al.*, 2012). Care should be taken on interpretation of these comparisons as *in vitro* conditions can differ largely between studies and different buffer components can greatly affect K_d values (Lohman and von Hippel, 1986; Hart *et al.*, 1999). For example, in this study, the addition of 50 mM NaCl resulted in five-fold difference of K_d . Furthermore, the fluorescence-based techniques used in this study involve the modification of DNA and specifically by addition of the 5'-FAM fluorophores, which can impact thermodynamics of the intermolecular interaction (Szabó *et al.*, 2018). Therefore, caution should be applied regarding the absolute values of DNA binding affinities and kinetic parameters, when interpreting the results. Nevertheless, these are efficient tools in probing a protein-DNA interaction to allow comparisons between, for example, different DNA ligands within the same buffer conditions.

Bypassing the need for DNA modification, AUC-SE was used to monitor DNA at an absorbance of 260 nm, distinct from protein interference at 280 nm. SE analysis provided evidence for untagged protein-DNA interaction, via the increase in calculated molecular mass. However, SE analysis and interpretation is more complicated than either FP or EMSA and reliable data was only obtained

under very particular conditions (for example, low concentrations of protein, addition of DTT). The stability of AtIPK1 was problematic in establishing SE profiles in preliminary experiments and further analyses were required to find a buffer, in which AtIPK1 remained stable for 24 h with additives that did not have a deleterious effect on AUC results. Additionally, it was later found that mixed populations of annealed and non-annealed 12-mer DNA were present in our SE experimental set up, which complicate our analyses. Nevertheless, these results suggest that AtIPK1 binds DNA in a 1:1 ratio. The oligomeric state of AtIPK1 binding DNA could be further analysed using AUC-SV, which typically takes only 4-6 h per run, and therefore does not rely on the longer-term stability of protein (Howlett *et al.*, 2006). However, SV analysis is not representative of monomer vs dimer distribution at equilibrium; and the problem of multiple configurations of ss vs dsDNA binding AtIPK1 will still present difficulties for data interpretation. Therefore, purification of dsDNA after the annealing protocol to ensure no ssDNA contamination is present would be required for future AUC analysis.

2.4.2 Non-specific DNA binding of AtIPK1

Non-specific DNA binding vs specific DNA binding was explored by limiting non-specific interactions by the addition of NaCl in binding buffers and assessing three different oligomers (FAM-AT 40mer, FAM-GC 40mer, and FAM-12mer) for binding preference by AtIPK1. With randomized double-stranded FAM-oligomer (GC 40mer), AtIPK1-DNA binding was observed in conditions with up to 100 mM NaCl. However, in FP assays at this higher NaCl concentration, full saturation of binding was not achieved despite AtIPK1 titrations up to 10 μ M, indicating a strong dependence of the AtIPK1-DNA interaction on electrostatic charges (Privalov *et al.*, 2011). Electrostatic interactions predominantly facilitate non-specific DNA binding, where large areas of positive charges on the protein surface attract the negatively charged phosphate backbone of DNA (von Hippel and Berg, 1986; Boyer, 2002). In contrast, specific DNA binding arises from hydrogen bonds and water-mediated interactions between protein side chains and nucleotide bases (Boyer, 2002). Without confirming specific DNA binding using

a high-throughput method such as systematic evolution of ligands by exponential enrichment [SELEX] (Chai *et al.*, 2011) or protein binding microarrays (Berger and Bulyk, 2009), we cannot exclude that AtIPK1 may have specific DNA binding activity.

Non-specific DNA binding describes any protein-DNA interaction which does not occur between protein and the target consensus sequence (Afek *et al.*, 2014). However, within non-specific DNA binding, preferences for different or specific sequence motifs can be quantified, and may have implications *in vivo* (Afek *et al.*, 2014). In this study, AtIPK1 interacted with similar affinity towards AT or GC rich 40-nt oligomers, where preference towards AT oligomers was not statistically significant (K_d of approx. 370 nM compared to 420 nM). Therefore, no preference of binding could be identified relating to AT/GC composition. It is possible that non-specific binding occurs completely independent of nucleotide sequence and entirely based on the geometry of DNA (von Hippel and Berg, 1986; Afek and Lukatsky, 2012), as the case may be for chromatin remodelling proteins (Saha *et al.*, 2006).

The High Mobility Group (HMG) superfamily proteins are well characterized non-specific DNA binding proteins involved in DNA replication and regulation of inducible gene transcription (Štros *et al.*, 2007). One such protein, HMGAI, has been shown to crystallize with a 12mer (Fonfría-Subirós *et al.*, 2012), an oligonucleotide which was synthesized for use in this study. We encountered difficulty whilst annealing the 12mer, a structure prone to hair loop formations (Fonfría-Subirós *et al.*, 2012); thus, in experiments using this oligonucleotide there were populations of double-stranded and single-stranded DNA. This was clearly visualized using F-EMSA, which furthermore suggested that AtIPK1 has a preference towards ssDNA (Figure 2.11). To validate this apparent preference, further analysis of strength of binding assays to ssDNA vs dsDNA need to be assessed using purified ss or dsDNA to ensure homogeneity and thus allowing clear interpretation of results.

2.4.3 Evidence of IPK dimer formation

In this study, it was shown through AUC-SE analysis that AtIPK1 (at concentrations of 8-16 μM) forms a dimer in solution. This was not supported from gel filtration chromatography (Appendix Figure A2.2) where a single peak molecular mass was observed with a molecular mass of ~ 50 kDa, corresponding to expected monomeric mass. This contradiction may be explained by (1) the highly concentrated AtIPK1 sample used in AUC-SE analysis compared to the protein eluate from chromatography and (2) the buffer conditions used to examine native AtIPK1 in AUC-SE (20 mM HEPES pH 7.5, 1 mM MgCl_2 , 50 mM NaCl, 0.5 mM DTT) differing substantially from the size exclusion elution buffer (20 mM Tris-HCl, pH 7.5, 200 mM NaCl). Specifically, the change from 200 mM NaCl in size exclusion chromatography to 50 mM NaCl in AUC may allow dimer stabilisation through electrostatic interactions at the dimer interface, whereas at higher salt concentrations these interactions are prevented (Purohit, 2018). In addition, gel filtration analysis does not separate molecules by size alone and the protein may interact with the Sepharose column directly (Burgess, 2018). Validation of the AtIPK1 dimer in solution may involve: AUC-SV to find monomer-dimer dissociation rates (Liu *et al.*, 2011), dynamic light scattering experiments which may help identify conditions for a stable AtIPK1 dimer (Bauer and Schnapp, 2007) or Blue NATIVE-PAGE analysis (Eubel *et al.*, 2005).

When the AtIPK1 crystal structure was solved by González *et al.* (2010), two molecules were found in the asymmetric unit, however it was considered unlikely that the protein was a physiological dimer because of a lack of hydrogen bonding between the proteins. Often proteins crystallize in the most efficient packing lattice such that even proteins that do form dimers *in vitro* do not form that same dimer in a crystal lattice (Marianayagam *et al.*, 2004). To further investigate the dimer structure of AtIPK1, purification of homogenous dimer by optimising the protocol for dimerization conditions, followed by further attempts with crystallization trials would be required.

The dimerization of AtIPK1 was previously observed by the Brearley lab in FP experiments (Whitfield, unpublished), where Alexa Fluor® 488-tagged AtIPK1 appeared to dimerize with additions of greater than/equal to 1 μM untagged AtIPK1 (and more favourably with $>10 \mu\text{M}$ AtIPK1) in the presence of excess DNA; and to lesser extent in the presence of both DNA and InsP_5 . However, in the absence of DNA, formation of the tagged/untagged-AtIPK1 heterodimer was not observed with additions up to 100 μM untagged AtIPK1 (Whitfield, unpublished). It was therefore considered that the presence of DNA promotes AtIPK1 dimer formation, which may then facilitate sequence specific interactions (Kohler and Schepartz, 2001) when AtIPK1 is present at a threshold abundance for finetuned regulation of DNA binding activity (Buchler and Louis, 2008). This contrasts with the findings within our AUC-SE study which are not ambiguous in predicting that untagged AtIPK1 is a dimer in solution (under the tested conditions: with 8 μM and above of AtIPK1), and which furthermore suggest that AtIPK1 binds to DNA in a monomeric form at lower protein concentrations (over a range of 1 to 4 μM of AtIPK1). These data lead us to hypothesise that DNA binding activity may disrupt or prevent AtIPK1 homodimer formation; or that a threshold concentration of AtIPK1 is needed for dimerization to occur. The latter may be considered since we could not obtain high quality data for molecular mass calculations of the DNA-protein complex at higher protein concentrations, nor have we measured the mass of AtIPK1 in solution at lower protein concentrations. Within studies that monitor unmodified *AtIPK1* expression, details pertaining to subcellular localization and abundance are not available, although a growing number of studies report nuclear localization of GFP-tagged AtIPK1 (Whitfield, 2013; Kuo *et al.*, 2014; Adepoju *et al.*, 2019).

The apparent contradiction of the observations made from these comparative studies may be explained by the different conditions for binding assays, especially with the addition of salt and DTT in the AUC study which can prevent protein aggregation. It is also possible that the dimer formation observed over the time scale of FP assays may not be stable over the 48 h typical of AUC experiments. Another consideration would be the presence of the Alexa Fluor® 488 dye which

may interfere with interactions with untagged AtIPK1 even at high concentrations of untagged protein and may be disrupted when DNA is interacting with the protein. It would be interesting to determine whether there is a threshold abundance of AtIPK1 necessary for homodimer formation using untagged protein in AUC-SE experiments and whether those conditions may be physiologically relevant.

Physiological dimers may have roles in DNA binding, such as for leucine zipper and zinc finger transcription factors for which dimerization increases the specificity to recognition binding sites (Yesudhas *et al.*, 2017; Sloan *et al.*, 2020). However, without the cognate nucleotide sequence this hypothesis is difficult to test for AtIPK1. Other roles for protein homodimers include sensing protein levels and regulation, such as localized increase of enzyme activity and facilitation of other protein interactions (Marianayagam *et al.*, 2004). At the time of reporting, there is limited evidence for IPK physiological dimers in the literature. IPK structures that have been published in the PDB, sometimes constitute dimers within the asymmetric unit of the crystal lattice, such as HsITPK1 (Chamberlain *et al.*, 2007), however in many cases further analyses such as gel filtration, static light scattering (Chamberlain *et al.*, 2007) or AUC (González *et al.*, 2004) have indicated a monomeric protein state. In exception to this, the dimeric form of AtIPMK α , an inositol polyphosphate multikinase with 3-/5-/6-kinase activity was supported by gel filtration analysis (Endo-Streeter *et al.*, 2012), yet has not been investigated for functional or physiological roles. Future experiments to further investigate the quaternary structure of AtIPK1 *in vivo* may include co-immunoprecipitation experiments (Chou *et al.*, 2015), or fluorescence microscopy techniques such as spatial intensity distribution analysis (Godin *et al.*, 2011) or FRET spectrometry (Raicu and Singh, 2013).

2.4.4 Towards *in vivo* evidence

As a proof of concept, under manipulated experimental conditions, AtIPK1 binds to dsDNA and ssDNA. However, *in vitro* experiments are very limited in applicability under physiological constraints. Despite the ability to recreate pH

or salt physiological conditions, the complexity of biological systems, including the entirety of the cellular contents and compartmental environments, are not accounted for. To explore whether DNA binding activity of AtIPK1 was observable *in planta*, FLIM-FRET was attempted. Following the protocol outlined by Camborde *et al.* (2017), we obtained FRET efficiencies similar to those reported in the paper for the DNA binding protein ArCRN13 (~26%, although not statistically validated, and 23%, $P < 0.0001$ respectively), however we calculated a much higher efficiency for the GFP negative control, 13.3% (not statistically validated) compared to the reported 1.6% $P < 0.5$ (Ramirez-Garcés *et al.*, 2016). Without a reliable negative control, it is difficult to interpret positive results; although at face value our results for AtIPK1-GFP do not contradict the hypothesis that AtIPK1 DNA binding is physiologically relevant.

Although the mean τ obtained for non-treated GFP of ~2.3 ns is not far off those reported in the literature [2.25 ns] (Ramirez-Garcés *et al.*, 2016), we observed a high degree of variability within single leaf disc samples leading to large error. GFP τ measurements can suffer interference from autofluorescence which may shorten GFP τ (Zhou *et al.*, 2005), a common disadvantage when working with plant tissues (Berg and Beachy, 2008). This is likely due to chlorophyll which localizes to chloroplasts close to nuclei (Roshchina, 2012; Camborde *et al.*, 2017); this fluorophore is excited by UV, blue or green light and has strong emission of red light (Lamb *et al.*, 2018). Modification of tissue preparation may help reduce autofluorescence interferences (Camborde *et al.*, 2017). When treated with SO, we observed a general decrease in GFP τ against expectation (~2.0 ns), which could be interpreted as GFP interaction with DNA however it has been shown that this is not the case (Ramirez-Garcés *et al.*, 2016; Camborde *et al.*, 2017). This result may be a by-product of error accumulated from τ measurements, and is not helped by the low sample size, $n = 9$, from which the mean was calculated.

Within this study small sample size was a limitation to data analyses, especially in the case of SO treatment where a lack of overlapping areas of GFP and SO within the leaf disc was a major bottleneck. Small sample sizes can lead to errors

in statistical analyses, as the effect size is inflated leading to false positive or negative rejections of the null hypothesis (Serdar *et al.*, 2021). Different solution infiltration methods were trialled on individual leaf discs, such as vacuum desiccation vs. syringe infiltration, in attempts to improve the area of SO permeation, but significant improvements were not seen. The required time to capture τ data was another factor contributing to small sample size. It was difficult to get through 20 individual nuclei FLIM measurements in a single day, whereas for more robust statistical analyses sample size recommendations were 30 measurements per construct per condition for a single experiment (Camborde *et al.*, 2017). Due to the time requirements and subsequent high cost, repeating the experiments to obtain sufficient sample size was unfeasible.

Future work to explore whether AtIPK1 binds to DNA *in planta* may involve using ChIP-seq (Zhu *et al.*, 2012), where plant tissue extracts and high throughput sequencing data can capture protein-DNA interactions and aid identification of potential DNA binding motifs (Chen *et al.*, 2017).

2.5 Conclusion

In summary, we have used fluorescence-based assays to explore quantitatively and qualitatively the non-specific binding activity between AtIPK1 and DNA. Although the DNA binding activity observed within this study produced K_d values which indicate physiological relevance, the salt conditions under which these were obtained do not reflect physiological conditions. Under the tested conditions with unlabelled protein and DNA, AtIPK1 appears to bind DNA as a monomer in a 1:1 ratio. However, we also observed AtIPK1 as a homodimer in solution at high protein concentrations ($\geq 8 \mu\text{M}$), which may or may not have physiological relevance. Preliminary steps *in planta* were indicative that AtIPK1-DNA binding may be a physiological phenomenon, however the FLIM-FRET methodology did not yield data sets for more reliable analyses and future work is needed to explore this question further.

An additional limitation of this current investigation is the absence of a target sequence for DNA binding activity. Many intensive protein-DNA *in vitro* studies have physiological target sequences, such as promoter regions, identified which would then be implied in further investigations for a particular cellular role. Despite limitations of individual approaches, taken together these data show that AtIPK1 interacts with DNA for all methods utilised. These findings are entirely novel in plant IPKs, with no interaction with DNA having been reported previously and offer an exciting insight into the potentially more complex role of such proteins.

Finally, it is worth mentioning the nuclear context of yeast and human IPK. The work of York (York et al. 1999) established a nuclear envelope pathway of InsP₆ synthesis comprising, PLC, IPK2 and IPK1 in yeast. The same pathway and nuclear localization is widely reported to exist in animals, though recent work (Desfougères *et al.*, 2019) suggests that other pathways may operate. Nevertheless, there is consensus that IPK1 catalyzes the final step of InsP₆ synthesis in plants, yeast and animals. It is interesting, therefore, that Brehm *et al* (2007) have speculated that human Ipk1 possesses a putative H2C2H2 zinc finger. Combined mutation of all the first four residues to serine elicited a shift to a predominantly cytosolic localization of a C-terminal eGFP fusion. Mutation of neighbouring lysine residues to alanine showed a similar but more pronounced effect. This region of human Ipk1 comprises either a nuclear import signal or a nuclear retention signal, though the mechanism by which this is achieved is unclear (Brehm *et al.*, 2007). Moreover in a later study, the same group (Brehm *et al.*, 2013) have attributed non-catalytic nucleolar function to human IPK1, specifically, interaction with proteins that regulate rRNA synthesis. Clearly, a nucleotide-related function for plant IPK1 has wider precedent.

3 AtIPK1 structure and mutagenesis

3.1 Introduction

In the previous chapter, it was established that AtIPK1 interacts with dsDNA and ssDNA *in vitro* and preliminary FLIM-FRET results *in planta* may suggest that this interaction is physiologically relevant. DNA binding proteins are generally characterized by their physiological action, target sequence and mechanism of binding. DNA binding proteins interact with target DNA through their tertiary/quaternary structures via DNA binding domains or binding motifs. The structure of AtIPK1 has previously been solved and is available for examination, however DNA binding activity has not been predicted by annotative software such as InterProScan. In this chapter, the mode of DNA binding of AtIPK1 was investigated from a structural perspective, examining potential DNA binding motifs facilitated by computational tools and inhibitor screening.

3.1.1 Structural characterization of AtIPK1

The first crystal structures of IPK1 from *Arabidopsis thaliana* were solved by González *et al* in 2010, with its substrate and products. In these early models, it was confirmed that AtIPK1 is structurally similar to the protein kinases (PKs) and belongs to the inositol phosphate kinase (IPK) family (González *et al.*, 2010), exhibiting the characteristic two lobes connected by a hinge, and particularly an N-terminal lobe $\alpha+\beta$ fold (Figure 3.1). Unlike other IPK subfamilies, IPK1 has an extensive C-terminal inositol binding pocket (Figure 3.1), referred to as the C-IP lobe, which encompasses almost half the protein (González *et al.*, 2010).

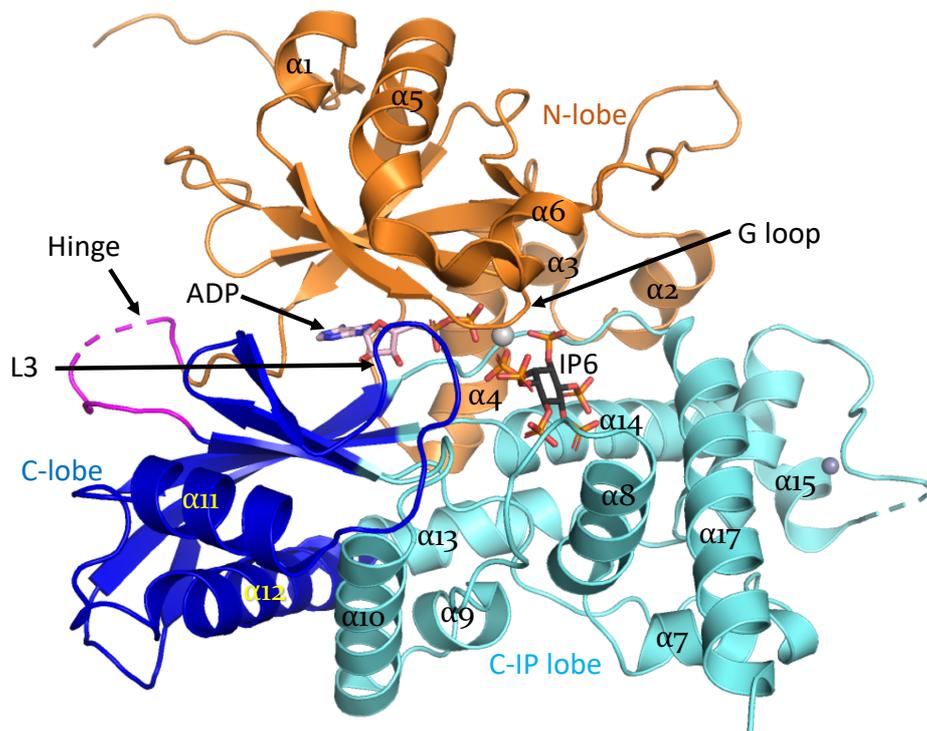


Figure 3.1 A 2.02 Å cartoon representation of AtIPK1 (PDB 6FJK) in complex with InsP₆ and ADP (represented as sticks), highlighting: N-term domain (orange), C-term (blue) and C-IP lobe (cyan).

3.1.2 AtIPK1 substrate specificity and conformation changes

Early AtIPK1 structure models in complex with ligands, in combination with site-directed mutagenesis (SDM) and biochemical enzyme assays were used to identify critical residues for substrate recognition (González *et al.*, 2010; Baños-Sanz *et al.*, 2012a; Gosein and Miller, 2013). InsP discrimination to achieve phosphorylation of axial 2-OH of the *myo* inositol ring is accomplished by tight co-ordination of all phosphate groups of the InsP₅ substrate, completed by the meeting of the N- and C-IP lobe (González *et al.*, 2010). Three phosphates are solely co-ordinated by C-IP residues: where K411, R415 and W419 co-ordinate P4, K170 (and through water H196 and R192) co-ordinate P5 and K170, K200 and N239 co-ordinate P6 (González *et al.*, 2010). In contrast, the N-lobe alone interacts with P1 via two hydrogen bonds with R130, from $\alpha 6$ (González *et al.*, 2010). P3 is held in place between lobes by C-IP K411 and N-term R45 and G loop main chain (González *et al.*, 2010). Lastly, the 2-OH acceptor is co-ordinated by the catalytically critical K168 and D368 (González *et al.*, 2010; Gosein and Miller, 2013)

Furthermore, AtIPK1 conformational changes were observed in crystallography models utilizing SDM (Baños-Sanz *et al.*, 2012a), enabling a model of enzyme activation to be proposed (Gosein and Miller, 2013). In the complete absence of substrate ligands, the N-lobe is more disordered and becomes unstable, as independently shown using limited proteolysis (Gosein *et al.*, 2012); therefore the production of apo AtIPK1 crystals, which require the formation of a crystal lattice, is challenging. However, an apo structure was eventually solved with the catalytically active W129A 'clasp' mutant (Baños-Sanz *et al.*, 2012a). The solved structure of the W129A mutant is emblematic of the open conformation of AtIPK1, where a more open cleft is formed between the C- and N-terminal domains (Figure 3.2). Upon nucleotide and InsP binding large changes are observed, the so-called half-open and closed conformations respectively, which is understood to take place through two rigid domain movements (Baños-Sanz *et al.*, 2012a). In an open conformation, L3 is mobile and acts as a lid to allow the entrance of ATP. When nucleotide substrate is bound, the G-loop residues coordinate the nucleotide phosphates (P_{β} and P_{γ}) via E18 and N21 respectively. In the closed state, the active site is enclosed by a fully formed $\alpha 6$ and L3, like a clasp. Formation of the clasp requires disruption of E255-K200 ion pair that link L3 and the C-IP lobe in open conformations (Baños-Sanz *et al.*, 2012a).

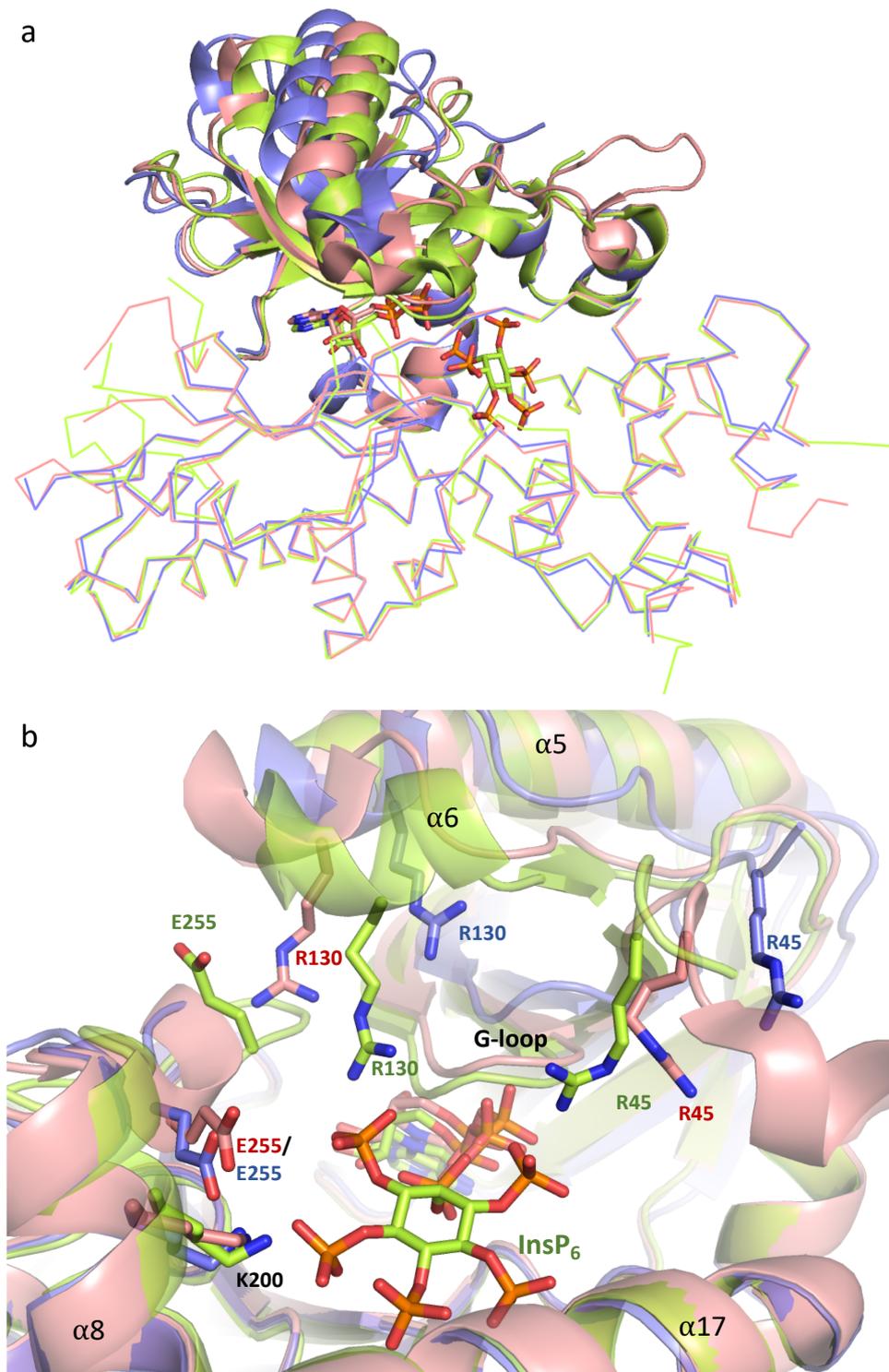


Figure 3.2 Superimposed models of the three conformations of AtIPK1: Showing open (slate blue), half-closed (pink) and closed (yellow) states (PDB entries: 4AXC, 4AXE and 2XAM respectively). Alignments of the entire structure are shown in (a) with flexible N-lobe also highlighted in cartoon representation, and a zoomed in view of the active site is shown in (b). Residues for InsP binding and stabilizing different conformations are highlighted as sticks (coloured by element). Text is coloured according to model and black text applies to all models.

3.1.3 Putative DNA binding motifs of AtIPK1

Other than elucidation of kinase activity, the solving of the AtIPK1 structure revealed the presence of an exposed zinc ion coordinated by H320, C330, C333, and H346 within the C-IP lobe (González *et al.*, 2010). These co-ordinate residues are conserved among plant IPKs presenting a non-canonical zinc finger (ZF) motif, H-X₉-C-X₂-C-X₁₂₋₁₅-H [Figure 3.3] (Whitfield, 2013). ZF motifs are small and functionally diverse domains, which may participate in enzymatic activity, protein structural stability or the binding interactions of proteins, DNA or ligands (Krishna *et al.*, 2003). To better understand the function of the AtIPK1 ZF, co-ordinating cysteines were the target of mutagenesis [e.g. C330A, C333A and C330A/C333A] (Whitfield, 2013). When expressed transiently with an N-terminal GFP tag in *Nicotiana benthamiana*, the AtIPK1 cysteine mutants showed altered compartmentalisation, wherein nuclear localization of single mutant GFP-AtIPK1 was not detected and was found only in the cytoplasm. The double cysteine mutant was detected with high intensity in the nucleus in a granular pattern suggesting aggregation of the protein, whereas it was largely absent from the cytoplasm (Whitfield, 2013). This suggested that the overexpressed double mutant was incorrectly folded and protein degradation machinery was not able to process the proteins for degradation as rapidly as it was being produced. Attempted expression of these mutants in an *E. coli* system resulted in the production of insoluble protein, preventing biochemical characterization and perhaps suggesting that the ZF is merely a structural feature of AtIPK1 (Whitfield, 2013), supporting the GFP microscopy findings discussed above.

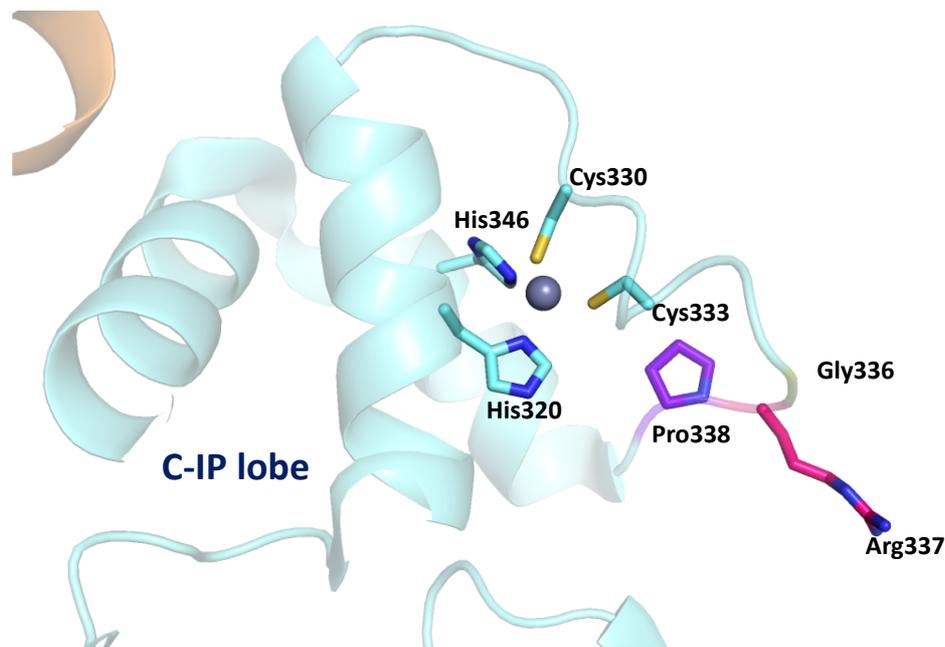


Figure 3.3 An expanded view of the AtIPK1 putative DNA binding motif, showing important residues as sticks (coloured by element) including the ‘GRP’ tripeptide: glycine-336 (green), arginine-337 (hot pink) and proline-338 (purple); and zinc co-ordinating residues. The protein chain is represented as a cartoon (displayed with 0.5 transparency) and the zinc ion (grey) as a sphere.

Within the ZF region, a flexible loop on the surface of AtIPK1 offers an alternative potential DNA binding feature: a derivative of the AT hook motif (PRGRP), specifically the GRP tripeptide (Figure 3.3). The GRP tripeptide was first described in high mobility group non-histone chromosomal protein HMG-I(Y) (Aravind and Landsman, 1998), and has been demonstrated as sufficient itself to bind DNA (Reeves and Nissan, 1990). This small motif is conserved in all *Arabidopsis thaliana* ecotypes (within a set of ecotypes which have been fully genotyped), with a single nucleotide polymorphism where R337 is replaced with K337 in approximately 50% of accessions (Whitfield, unpublished). Given that arginine and lysine are both amino acids with positively charged side chains it would be anticipated that the change in residue would have minimal effect on the properties of the tripeptide. Interestingly, it was observed in *Arabidopsis lyrata* that the sequence is not conserved, with the residues SGP in place of GRP. This offers a natural sequence variation, where changes to residue properties from GRP to SGP may result in an inability of *A.lyrata* IPK1 to bind DNA.

3.1.4 Aims of experiments

In this continued study, we set out to determine the mechanism of the AtIPK1-DNA interaction. Within the last decade, the structure of AtIPK1 has been studied by two research groups, with attention to the nucleotide and InsP binding pocket, to elucidate the mechanism of enzyme activation and product formation (González *et al.*, 2010; Baños-Sanz *et al.*, 2012a; Gosein and Miller, 2013). Here, we attempt to solve the AtIPK1-DNA complex and reveal the DNA binding site by conducting crystallization screening of AtIPK1 with annealed dodecamer, utilizing the available expertise and equipment from the UEA School of Biological Sciences. In the absence of the complex structure, *in silico* analyses were used to identify and explore potential DNA binding residues.

An inspection of the AtIPK1 structure reveals two putative DNA binding motifs, which have not been fully characterized, including: the non-canonical ZF (Whitfield, 2013) and the GRP tripeptide. These disparate motifs were observed within the same region of the AtIPK1 C-IP lobe but would characteristically bind DNA quite differently. For example, ZF transcription factors are typically major groove binding proteins with high sequence specificity (Krishna *et al.*, 2003); whilst, AT hook proteins are minor groove binding proteins often involved in non-specific DNA binding (Aravind and Landsman, 1998). Whilst production of AtIPK1 mutants with substitutions of cysteines co-ordinating the zinc ion were unsuccessful due to protein insolubility (Whitfield, 2013), the focus of this study will be to examine the GRP tripeptide as a potential DNA binding site.

The GRP motif was investigated using site-directed mutagenesis (SDM) to obtain three mutants: p.G335_P337del (referred to as Del(GRP)), p.R336K (referred to as GKP) and p.G335S/R336G (referred to as SGP). Fluorescence polarization and fluorescence-based EMSAs were used to obtain quantifiable DNA binding activity (via EC50 values) and visual confirmations, in order to compare with wildtype AtIPK1 and examine: (1) the effect of removing this potential DNA binding motif, (2) the impact of the GRP polymorphisms observed amongst *Arabidopsis* ecotypes and (3) the Ser336/Gly337 variant of *A. lyrata*.

ZFs are amongst classic examples of DNA binding motifs common in plant transcription factors (Riechmann *et al.*, 2000), although they may also be involved in protein-protein interactions, including homodimerization (Jauch *et al.*, 2003). We have demonstrated that AtIPK1 can dimerize *in vitro* (see Chapter 2) thus *in silico* modelling was used to explore potential dimer sites. Dimer models were also inspected for potential co-operative DNA binding sites.

Lastly inhibitor assays using FP displacement assays were used to explore whether substrate or product binding of AtIPK1 may affect DNA binding activity, and whether netropsin, a minor groove binding ligand, may prevent DNA binding activity.

3.2 Materials and methods

3.2.1 Crystallography trials with AtIPK1 and dsDNA

3.2.1.1 Crystallography screens with AtIPK1

Freshly purified AtIPK1 (10 mg/mL) was incubated with annealed dodecamer oligonucleotide [12mer] (5'-CGAATTAATTTCG-3') at monomer (1:1.2) and dimer (1:2.4) molecular ratios of DNA:protein for 1 h on ice, and centrifuged at 13,000 rpm for 10 min in preparation for crystallization trials.

Crystallization of AtIPK1 was attempted using sitting drop vapour diffusion with six commercial screens from Molecular Dimensions: (Structure™ 1&2 Eco Screen (Jancarik and Kim, 1991), JCSG-plus™ Eco Screen (Collins *et al.*, 2005), PACT premier™ Eco Screen (Newman *et al.*, 2005), Morpheus® Screen (Gorrec, 2009), MIDAS™ Screen (Grimm *et al.*, 2010)). An OryxNano robot (Douglas Instruments Ltd.) was used to set 0.5 µl drops of protein and precipitant (1:1) in 96-well 2-drop MRC crystallization plates (Swissci), with a 50 µl reservoir of screen buffer. A pair of drops were set for each precipitant, with the monomer or dimer protein-DNA ratio. Each screen was conducted at 4 and 16 °C. Screen plates were checked every day using SZX9 Stereo Microscope (Olympus) for one week and subsequently twice a week for a month and finally once a week every month.

After two weeks, crystals were found growing in three conditions. In the monomer DNA-protein ratio, crystals grew at 16 °C in 0.2 M zinc acetate dihydrate, pH 6.5, 0.1 M sodium cacodylate, 18% w/v PEG 8000, and also at 4 °C with 2 M sodium chloride, 10% w/v PEG 6000. The final growth condition was in a dimer ratio at 16 °C with 0.1 M BisTris, pH 6.0, 25% SOKALAN CP42.

3.2.1.2 Crystal optimization, harvest and data collection

For optimisation, 96-well 2-drop MRC crystallization plates were set-up in the conditions described above using 1 µL droplets of protein and precipitant (1:1). The highest quality crystals were harvested with mounted cryo loops and soaked, using respective well precipitants with 20-30% ethylene glycol or glycerol as a cryoprotectant, before flash freezing with liquid nitrogen. Crystal pucks were

couriered in a liquid nitrogen dewar to Diamond Light Source (DLS: Oxford, UK) for remote data collection, using the i03 and i04 beamlines.

3.2.1.3 Data processing and refinement

The X-ray diffraction data was treated automatically by the Diamond processing pipeline and subsequently processed using the Phenix v1.12 (Adams *et al.*, 2010) system. Data quality was assessed using Phenix Xtriage, and if necessary, image files were integrated using only partial data to improve the data set. Molecular replacement was performed using Phenix Phaser with AtIPK1 PDB 6FJK as a search model for the AtIPK1-ADP structure and AtIPK1 W129A AMPNP-bound model PDB 4AXD (Baños-Sanz *et al.*, 2012a) for the apo protein. Phenix ReadySet was used to restrain the CHHC co-ordinating the zinc ion. Structures were improved with iterative cycles of model building in Coot v0.8.9.1 (Emsley *et al.*, 2010) and refinement with Phenix Refine to produce final models. In the ADP structure, an omit map verifying the presence of ADP was produced using simulated annealing in Phenix Refine.

Images of crystal structures were prepared with the PyMOL Molecular Graphics System v2.3.5 (Schrödinger, LLC). This software was also used to make alignments for comparisons between AtIPK1 structures and the measurement feature was used to view molecular distances.

3.2.2 SDM of AtIPK1

3.2.2.1 Primer design and PCR

XLI-blue with *pEAQ-HT-GFP[CaMV 35S-AtIPK1]* stocks were provided by Hayley Whitfield. These constructs were purified and subsequently heat-shock transformed (as described in 2.2.1.3) into competent *E. coli* Stellar and NEB stable cell lines, and glycerol stocks were made (as described in 2.2.1.4). Plasmids were propagated and isolated (as described in 2.2.1.1) to be used as template DNA for SDM.

To generate mutants of the AtIPK1 ‘GRP’ motif [referred to as GKP, SGP and del(GRP)] in *pEAQ-HT-GFP*, SDM PCR and primer design was performed following an adaptation of the QuikChange protocol (Liu and Naismith, 2008; Xia *et al.*, 2015) with Phusion™ high fidelity polymerase (NEB) and overlapping primer pairs (Table 3.1). For each mutant, a 25 µL reaction volume (10-20 ng/µL template, 0.4 µM of each primer, 100 µM each dNTP, 4% DMSO and 0.2 µL enzyme) was prepared in 1x HF buffer. Gradient PCRs were carried out for each primer pair from 50 to 72 °C in order to find the most efficient annealing temperature, and DNA gel electrophoresis was performed for analysis (see methods 2.2.1.2). A thermocycler was used for initial incubation of 98 °C for 3 min denaturing step, 25 cycles of 98 °C for 30 s, a gradient was applied 50 to 72 °C for 1 min annealing and 68 °C for 9 min 36 s extension, followed by a final extension at 68 °C for 10 min.

Table 3-1 Primers for SDM of AtIPK1. The nucleotides encoding the GRP motif are underlined, changed nucleotides in bold, deleted region presented as ‘|’ and the non-overlapping region is italicized.

Primer name	Primer sequence 5' → 3'	Overlap (bp)	GC (%)	Tm (°C)
GKP F	GTAAAGAAG <u>GTAAGCC</u> ACTGGAGGC <i>GGAATTG</i>	23	50	56
GKP R	CTCCAGTGGCTTACCTTCTTTACATA <i>TAGGAC</i>		44	
SGP F	GTAAAGAA <u>AGTGGGCC</u> ACTGGAGG <i>CGGAATTG</i>	23	53	56
SGP R	CTCCAGTGGCC <u>CACTTT</u> CTTTACAT <i>ATAGGAC</i>		47	
del(GRP) F	GTAAAGAA CTGGAGGCGGAATTGT <i>CTCTACATGCTTTAC</i>	19	44	54
del(GRP) R	CCGCCTCCAG TTCTTTACATATAGG <i>ACAAGGCTGGTTG</i>		50	

Products from optimised 20 µL PCR reactions were digested with 0.5 µL *DpnI* (Promega) incubated at 37 °C for 2 h and then purified using the Wizard® SV Gel and PCR Clean-Up System (Promega), and heat-shock transformed into

competent Stellar strains for plasmid propagation. Samples of the recovered plasmids were sent for sequencing using the Mix2Seq kit (Eurofins) with pOPINF F primer (Table 3.2) as per manufacturer’s instruction. These data were analysed using 4Peaks v1.8 (Nucleobytes) for quality and checked with Clustal Omega (Madeira *et al.*, 2019) for fidelity to the wildtype sequence and the presence of the desired mutations.

3.2.2.2 In-fusion cloning AtIPK1 mutants into pOPINF

To prepare for cloning into pOPINF, successful SDM plasmids of GKP, SGP and Del(GRP), generated as described in 1.1.3.1, were used to amplify the *AtIPK1* insert with pOPINF adapter 5’ and 3’ ends (Table 3.2). This was performed using Phusion™ (Thermofisher) high fidelity polymerase as per manufacturer’s instruction in a 20 µL reaction volume (10-20 ng/µL template, 0.5 µM of each primer, 200 µM each dNTP and 0.2 µL enzyme). A thermocycler was used for an initial incubation of 98 °C for 30 s denaturing step, then 30 cycles of 98 °C for 10 s, 47 °C for 20 s annealing and 72 °C for 45 s extension, followed by a final extension at 72 °C for 10 min.

Table 3-2 AtIPK1 primers for In-fusion into pOPINF, where underlined sequences are adaptor regions for integration into pOPINF and Tm calculated using Tm Calculator (NEB).

Primer name	Primer sequence 5’ → 3’	Adapter (bp)	GC (%)	Tm (°C)
pOPINF F	<u>aagttctgtttcagggcccgatggagatgattttggagg</u> agaa	20	47	75
pOPINF R	<u>atggtctagaaagcttttagctgtgggaaggtttt</u> g	18	43	69

Double restriction digest was performed on pOPINF (~1 µg) with 0.5 µL each of *KpnI* and *HindIII* (Promega), 0.2 µL BSA and 1x MULTI-CORE buffer provided in a 20 µL reaction volume as per manufacturer’s instruction. The reactions were incubated at 37 °C for 3 h and the products were run on 0.7% DNA gel. The linearized plasmid was excised under UV and extracted using the Wizard® SV Gel and PCR Clean-Up System (Promega).

Ligation independent cloning using the In-fusion (Clontech) system was performed according to manufacturer's instruction. The amounts of respective linear plasmid and insert for the SDM mutants used in the In-fusion reaction were determined by running samples of both on an 0.7% agarose gel. A positive control reaction of linear pUC19 with provided insert was performed in tandem. After a 15 min incubation at 50 °C the entire reaction volume was added to competent stellar cells (Clontech) and a heat-shock transformation protocol was followed according to manufacturer's instruction. Transformation reactions were spread onto LB agar plates containing 20 mM IPTG, 80 µg/ml X-Gal (5-Bromo-4-Chloro-3-Indolyl β-D-Galactopyranoside) and 100 µg/ml ampicillin for blue-white screening selection. Single white colonies were picked and grown in 10 mL LB with 100 µg/ml ampicillin, to make new glycerol stocks and to propagate for plasmid isolation. Plasmids were sent for sequencing with T7 and T7 term sequencing primers using Mix2Seq kit (Eurofins), which were analysed using ApE v2.0.49, 4peaks v1.8 (Nucleobytes) and Clustal Omega (Madeira *et al.*, 2019).

3.2.2.3 Transformation into competent *E. coli* strains

The SDM AtIPK1 mutant constructs were isolated using Wizard® Plus SV MiniPreps DNA Purification System (Promega), these were verified by Mix2Seq sequencing, and were used in transformation reactions into several competent cell lines: NEB stable competent (NEB), SHuffle® T7 Express (NEB), SHuffle® T7 Express LysY (NEB), BL21[DE3] (NEB), Rosetta[DE3] (Novagen), Rosetta2 (Novagen) and ArcticExpress™ RIL (Agilent Technologies) as described in methods section (2.1.1.3) for small scale expression trials.

3.2.3 Expression and purification of GRP mutants

3.2.3.1 Expression trials of AtIPK1 GRP mutants

Starter cultures of each SDM construct from every expression cell line where transformations were successful, were grown in 10 mL LB with 100 µg/ml ampicillin, overnight at 37 °C, 180 rpm. The next day, 1 mL of starter culture was used to inoculate 50 mL fresh LB (containing 100 µg/ml ampicillin) for each SDM construct. These cultures were grown at 37 °C, 180 rpm for 3-6 h and absorbance

readings at 600 nm (A600) of 1 mL at 600 nm were observed at regular time intervals against an LB media blank. Upon A600 readings of approx. 0.6, samples of 100 μ L were collected and prepared for SDS-PAGE analysis (see details in section 2.2.1.8) of non-induced cultures. The remaining cultures were induced with 0.2 or 0.4 mM IPTG and expressed overnight at rt, 180 rpm. Samples of 100 μ L for induced cultures were taken and prepared for SDS-PAGE analysis.

3.2.3.2 Solubility test of AtIPK1 SDM mutants

Expression cultures of AtIPK1 GRP mutants (GKP and Del(GRP)) from Rosetta, Rosetta2 and SHuffle[®] Express cells in 50 mL LB were pelleted after induction (described above), and cell pellets were resuspended in 2 mL lysis buffer (50 mM NaH₂PO₄ pH7.5, 300 mM NaCl, 20 mM imidazole, 0.5% triton), snap frozen and stored at -20 °C. Thawed cell suspensions were sonicated in seven sets of 10 s pulses with 30 s rest on ice between repetitions. The cell lysates were centrifuged at top speed for 20 min and the supernatant was carefully tipped off and stored at -4 °C. Samples of the pellet and supernatant were taken for SDS-PAGE analysis to assess protein in insoluble and soluble fractions respectively.

3.2.3.3 Large scale protein production of AtIPK1 GRP mutants

Large scale gene expression of SHuffle[®] T7 Express LysY (NEB) strains containing AtIPK1 SDM mutants [GKP, SGP and Del(GRP)] in pOPINF were conducted as per manufacturer's instructions, with 0.4 mM IPTG induction of 1 L cultures with overnight at 16 °C, 180 rpm. Cells were harvested as described in 2.2.1.5 and cell lysis was achieved using the French pressure cell as detailed in 2.2.1.6.

3.2.3.4 Purification of AtIPK1 GRP mutants

Purification steps for all mutants as described for wildtype AtIPK1 in section 2.2.1.7.

3.2.4 SDS-PAGE analysis

As described in 2.2.1.8 wherein additionally, purified wildtype and the three SDM mutants were compared by running samples side-by-side at 2 or 5 μ M.

3.2.5 HPLC assays

Activity assays were set-up in triplicate to measure the conversion of ATP to ADP with 250 or 500 nM protein (wildtype AtIPK1 and SDM mutants) and 50 μ M substrates (ATP and Ins(1,3,4,5,6)P₅) in 20 mM HEPES, pH 7.3, 1 mM MgCl₂. A negative control, for time point 0 min (i.e., reaction mixture without enzyme), was prepared similarly. The reactions were performed at rt in 25 μ L volumes for each protein and 5 μ L samples were taken at the following time points: 2, 4, 6, 8 and 10 min. The enzyme activity was quenched with an equal volume of 60 mM ammonium phosphate (pH 3.5); and samples were diluted 5x with ultrapure dH₂O. Injection volumes of 20 μ L were resolved by ion pair reverse-phase HPLC on a 200 mm x 4 mm i.d. Shandon Hypersil BDS C18 column, after (Caddick *et al.*, 2008). Nucleotides were eluted with mobile phase containing 5 mM tetrabutylammonium hydroxide in 50 mM NaH₂PO₄/acetonitrile (70/30), at a flow rate of 1 ml/min and baseline resolved within 10 min.

The proportions of nucleotide recovered as ADP and ATP were determined by monitoring the absorbance of the column eluate at 258 nm: the areas of the ADP and ATP peaks were integrated using the ChromNAV software (Jasco). Percentage of product formation was determined by using the nucleotide peak ratios. Data was plotted using GraphPad, where linear regression analysis was used (with $x=0$, $y=0$ as starting point) to calculate initial reaction rates. Images were edited using Microsoft PowerPoint.

3.2.6 FP assays

3.2.6.1 Wildtype AtIPK1 vs SDM mutants

FP assays to produce standard curves with wildtype and mutant AtIPK1 binding 2-FAM-InsP₅ and FAM-oligomers with 100 or 50 mM NaCl in the binding buffer, and all data analyses were carried out as described in section 2.2.3.

3.2.6.2 AtIPK1-DNA displacement assays

FP assays to displace the interaction between 5'-FAM-AT 40mer and wildtype AtIPK1 (at 2 nM and 500 nM respectively) were performed in 20 mM HEPES pH 7.5, 1 mM MgCl₂, 50 mM NaCl buffer using additions (0 to 10 μM) of the following ligands: ATP, ADP, Ins(1,3,4,5,6)P₅, InsP₆, netropsin and purpurogallin (Appendix 3). Each 100 μL reaction was incubated on ice for 1 h before aliquoting four replicates of 20 μL into each well of the 384 well Corning low volume black, flat bottomed, polystyrene plate. Polarization was measured at 25 °C using a Clariostar (BMG LabTech) plate reader with excitation at 485±5 nm and emission at 520±5 nm.

The raw data was used to calculate the mean and standard deviations of ≥ 3 separate measurements using Prism v6.0 (GraphPad). Non-linear regression analysis with variable slope model was used to produce standard curves and obtain IC₅₀ measurements.

3.2.7 F-EMSA with wildtype AtIPK1 vs SDM mutants

Wildtype AtIPK1 vs three SDM mutants were run on F-EMSA gel (as described in 2.2.2.3) at 1 and 5 μM with 2.5 nM FAM-12mer. The binding buffer conditions used were 20 mM HEPES pH 7.5, 1 mM MgCl₂, 0.05 mg/mL BSA, 50 mM NaCl and 1 mM EDTA.

3.2.8 *In silico* investigation of AtIPK1

The electrostatic potential from AtIPK1 structural models were examined in PyMOL using the APBS feature and the structure-based prediction server BINDup (Paz *et al.*, 2016) was used to predict positive binding surfaces of AtIPK1 and AtITPK4 which may bind DNA. Potential DNA binding residues were identified using DISPLAR (Tjong and Zhou, 2007), which considers physicochemical properties of adjacent residues from structural models and identifies likely DNA binding residues from highly conserved residues. Consurf-DB (Chorin *et al.*, 2020) was used to assess the highly conserved vs poorly conserved residues for AtIPK1, for reference. Finally, HDOCK (Yan *et al.*, 2017)

was used to generate models of the AtIPK1-DNA complex and identify potential DNA binding sites using a homology-based approach. Structural models of nucleotide dodecamer (12mer) from PDB ID 3UXW and wildtype AtIPK1 in different conformations were used as template for docking runs (including apo, ADP-bound AtIPK1 and ternary complexes: PDB 2XAM and 6FJK).

The AtIPK1 homodimer interface was investigated using the nucleotide bound wt AtIPK1 model for webservers: GalaxyHomomer (Baek *et al.*, 2017) and HDOCK. In the latter server, the C2 symmetry group was selected as a required parameter for modelling. Model alignments from GalaxyHomomer and HDOCK were performed in PyMOL v2.3.5 (Schrödinger, LLC).

All images were composed using PyMOL and edited in Microsoft Powerpoint.

3.3 Results

3.3.1 AtIPK1 co-purifies with nucleotide

In this work, it has been demonstrated that AtIPK1 binds FAM-labelled or unlabeled DNA *in vitro* (Section 2.4.1). To understand how the protein may bind DNA, its structure was investigated through X-ray crystallography. Crystallization of AtIPK1 with unlabeled dsDNA substrate was attempted using the sitting drop vapour diffusion method across six commercial screens. In initial screening with dimer and monomer ratios of protein to DNA, crystals grew under several growth conditions and individual crystals were harvested and sent to the synchrotron for data collections. However, no datasets containing the AtIPK1-DNA complex were obtained, instead crystals contained individual components (e.g., AtIPK1, DNA) or salt or diffracted poorly.

Despite no addition of nucleotide in the preparation for crystal screening, the first solved structure was of AtIPK1 in complex with ADP (Table 3.3; Figure 3.4), from growth conditions of: 0.1 M BisTris pH 6.0 and 25% SOKALAN CP42; and cryoprotectant of 30% ethylene glycol and 70% well solution. The presence of ADP was confirmed by generation of an omit map (Figure 3.4c). This discovery prompted the inclusion of a nucleotide stripping step during future purification of AtIPK1 (section 2.2.1.7), by means of a heparin column (Appendix Figure A2.2). Nucleotide contamination was further monitored with A260/A280 ratios on a nanodrop and was found to be approximately 0.6 for each prep indicating minimal contamination. Crystallization of AtIPK1 with DNA was thus attempted again using heparin purified protein and yielded crystals growing in a new condition at 4 °C in 2 M NaCl and 10% w/v PEG 6000. These crystals were subsequently harvested and sent for remote data collection, and a novel AtIPK1 apo structure was solved. However, the AtIPK1-DNA structure remained unobtainable from this trial. Considering the difficulties of solving protein crystal structures with DNA, and due to time limitations, this study was not taken any further.

Table 3-3 Crystal data collection, structure determination, and refinement.

	AtIPKI	AtIPKI-ADP
Wavelength	0.9762	0.9762
Resolution range^a	56.47 - 2.81 (2.911 - 2.81)	58.32 - 2.23 (2.31 - 2.23)
Space group	P 2 ₁ 2 ₁ 2	P 2 ₁ 2 ₁ 2
Unit cell	69.64 96.48 64.8 90 90 90	71.51 100.77 63.31 90 90 90
Total Reflections	21538 (2147)	44898 (4377)
Unique Reflections	10031 (1089)	22829 (2236)
Multiplicity	2.0 (2.0)	2.0 (2.0)
Completeness (%)	99.00 (98.97)	99.37 (98.76)
Mean I/sigma(I)	9.91 (1.47)	7.44 (1.30)
Wilson B-factor	80.61	39.2
R-merge	0.04823 (0.7019)	0.06317 (0.4393)
R-meas	0.06820 (0.9927)	0.08933 (0.6213)
R-pim	0.04823 (0.7019)	0.06317 (0.4393)
CC_{1/2}	0.998 (0.465)	0.995 (0.664)
CC*	1 (0.797)	0.999 (0.894)
Reflections used in refinement	11027 (1087)	22823 (2235)
Reflections used for R-free	536 (43)	1131 (141)
R-work	0.2739 (0.4607)	0.2020 (0.2677)
R-free	0.3035 (0.3842)	0.2740 (0.3417)
CC(work)	0.938 (0.446)	0.957 (0.821)
CC(free)	0.847 (0.72)	0.926 (0.747)
No. of non-hydrogen atoms	3179	3634
 macromolecules	3169	3452
 ligands	1	40
 solvent	9	142
Protein residues	398	436
RMS^b (bonds)	0.003	0.007
RMS (angles)	0.68	0.95
Rotamer outliers (%)	3.68	3.12
Clashscore	10.76	14.78
Average B-factor	98.59	43.4
 macromolecules	98.66	43.33
 ligands	139.09	48.79
 solvent	67.07	43.65
No. of TLS groups	7	
Ramachandran favoured (%)	89.69	93.55
Ramachandran allowed (%)	9.02	5.99
Ramachandran outliers (%)	1.29	0.46

^a values in parentheses are for highest-resolution shell

^b RMS = root mean squared

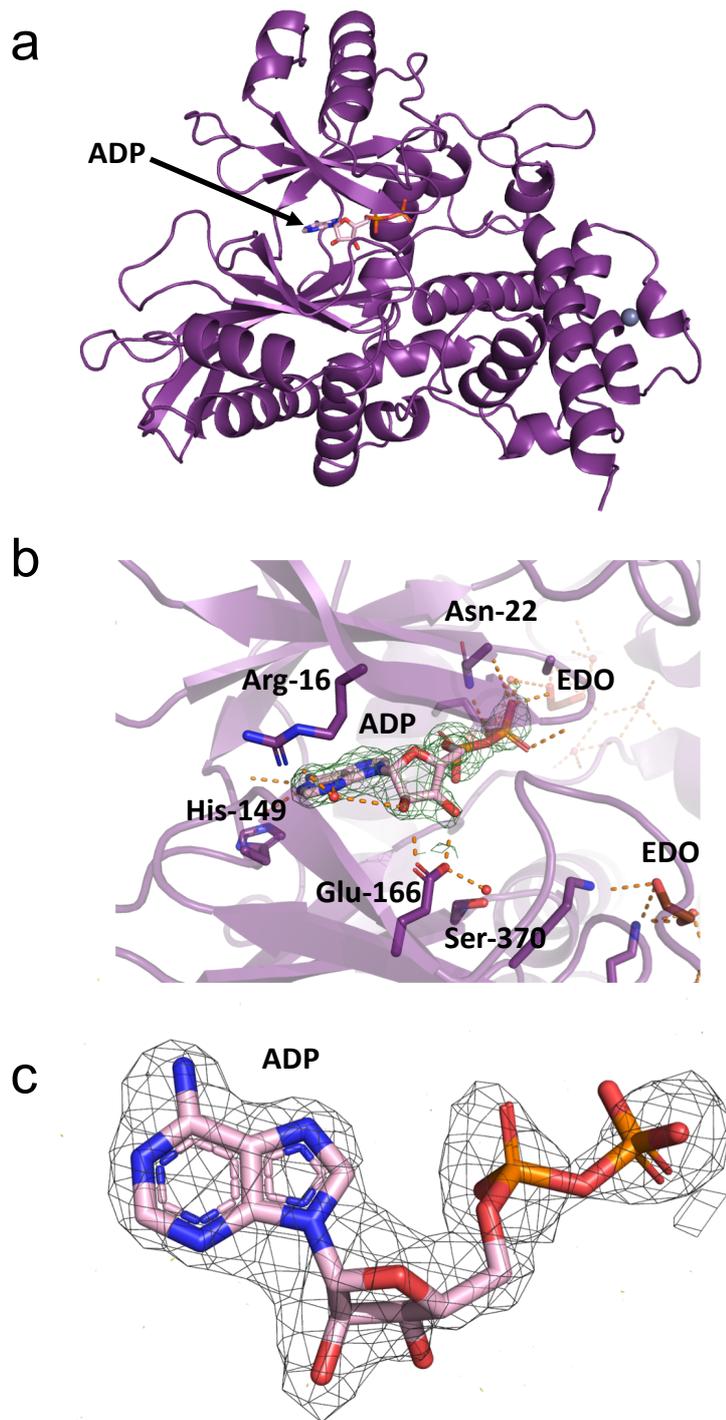


Figure 3.4 Cartoon representation of the 2.23 Å AtIPK1-ADP structure (purple). Ligands are shown as sticks coloured by element: ADP (pink), ethylene glycol (brown); and the zinc ion (grey) a sphere. (a) The overall structure of AtIPK1 in complex with ADP, showing ADP. (b) An expanded view of the AtIPK1 structure (chain transparency set to 0.5) where ADP is co-ordinated by residues (sticks, coloured by element), and a double difference electron density map is displayed as a green mesh over the ADP at a contour level of 1.2 r.m.s.d. (c) ADP with single difference density from an omit map with simulated annealing displayed at a contour level of 2.1 r.m.s.d.

3.3.2 Novel apo AtIPK1 structure

A 2.81 Å novel apo AtIPK1 structure was solved using molecular replacement with the nucleotide bound AtIPK1 W129A structure (PDB 4AXD) as search model, where no suitable solutions were found with the W129A apo structure (PDB 4AXC). The differences between apo and ADP bound structures solved in this study were highlighted using alignments in PyMOL (Figure 3.5). Overall, the structures were similar with RMSD of 0.764 (335 common atoms), and some features such as the L3 loop appearing somewhat displaced in the apo model RMSD of 2.315 (11 common atoms). Within the apo model, there are some poorly formed helices, such as $\alpha 5$ and $\alpha 6$ (similar to the ADP structure), as well as missing residues, such as at the 'hinge', which were difficult to model due to the quality of diffraction data i.e., no density. This lack of density is likely due to the absence of a bound substrate, meaning the hinge region is highly flexible since the N-term domain has not formed the residue contacts to keep it held in a closed position and thus stabilise the hinge region.

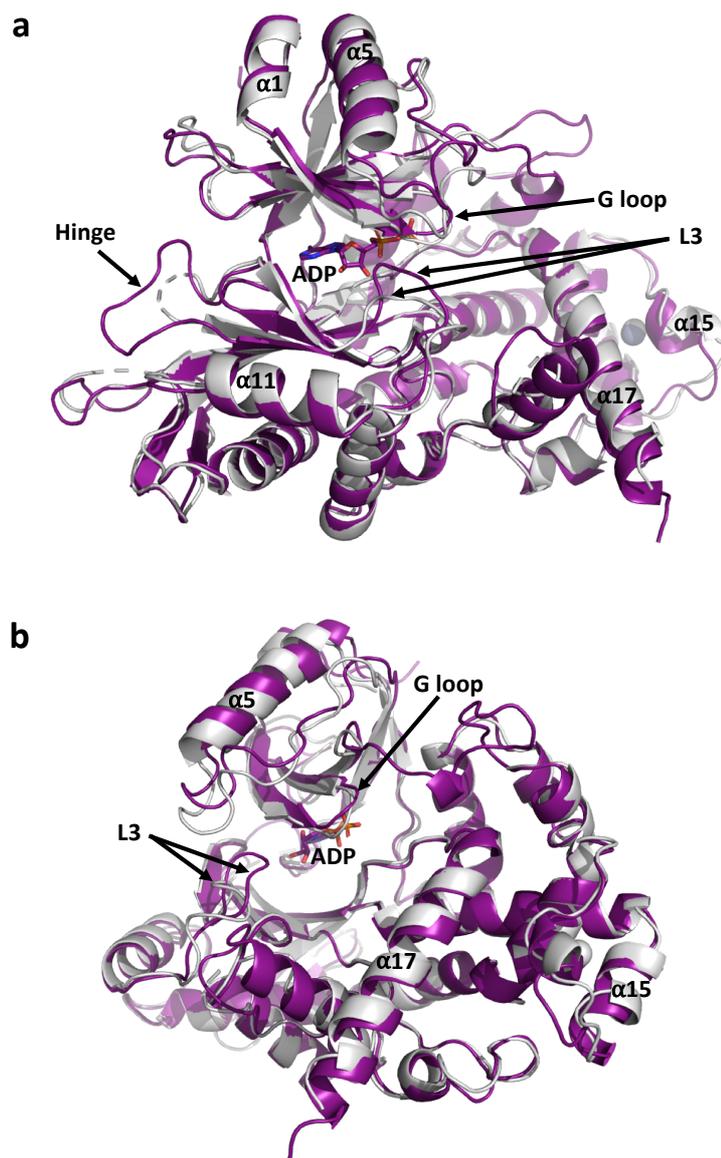


Figure 3.5 Two orientations of the superposition of wildtype apo AtIPK1 model (grey) against ADP bound AtIPK1 model (purple); with peptide presented by cartoon and ADP shown as sticks (coloured by element).

3.3.3 Comparison of AtIPK1 structure models

To allow comparisons between previously published structures in various conformations and the wildtype AtIPK1 structures solved here, further alignments were made in PyMOL. As expected, the wildtype AtIPK1-ADP complex was very similar to the previously reported ADP-bound W129A mutant (PDB 4AXE), in a half-open conformation, which resides within the same space group $P2_1 2_1 2$ and has overall RMSD of 0.288 (382 common atoms) [Appendix 4]. The largest difference between the two structures appears to be in the

configuration of L3, where there is no ion pair between K200 and E255, rather there appears to be stacking interactions between N-term W129 and C-term E255.

Interestingly, in whole structure alignments, the wildtype apo structure was equally similar to the half-open, ADP-bound structure [RMSD = 0.752 (346 common atoms)] and the 'open' conformation reported for the W129A apo structure [PDB 4AXC; RMSD = 0.7 (289 common atoms)] (Figure 3.6). To examine this further, N- and C-domains were looked at separately. The C-lobe, which contains the InsP binding pocket, was highly similar in comparisons with W129A ADP bound [RMSD = 0.473 (222 common atoms)] and W129A apo structures [RMSD = 0.475 (214 common atoms)]. Yet there appears to be a more emphatic movement of the N-lobe between the two apo structures [RMSD = 1.287 (123 common atoms)] compared to the nucleotide bound form 4AXE [RMSD = 0.939 (122 common atoms)], where the wildtype apo structure is a little more 'closed' than the W129A mutant, but less so than the ADP bound model (Figure 3.6c-d). In particular, the G-loop, which co-ordinates nucleotide phosphates via E18 and N21 (González *et al.*, 2010), and alpha helices -1, -5 and -6 appear to be rotationally shifted. On close inspection of symmetry mate interactions, the N-term lid domain residue T123 contacts a symmetry mate residue R178, present in both W129A apo structure and WT apo structure. This suggests that close symmetry mate contacts are not preventing the open form of the enzyme seen in the W129A apo structure.

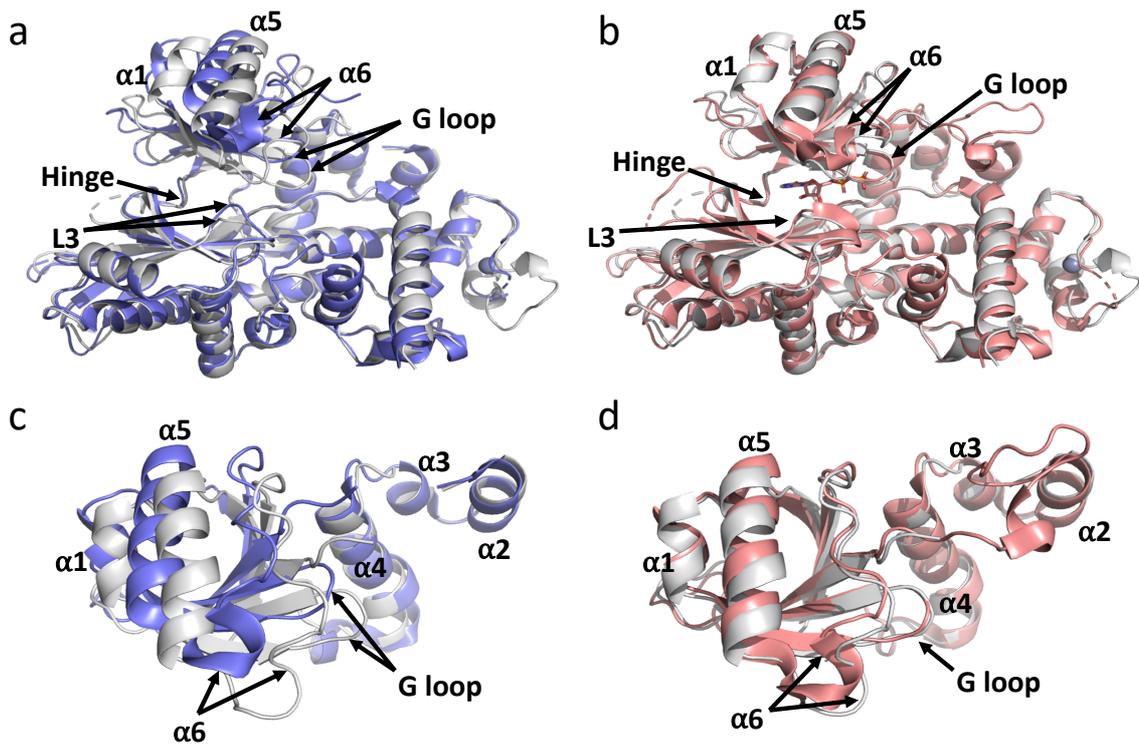


Figure 3.6 Superposition of native apo AtIPK1 model (grey) against the apo W129A model (a,c,e-f; blue; PDB 4AXC) and ADP-bound native AtIPK1 (b,d-f; pink; PDB 4AXE). Alignments of AtIPK1 monomer structures (a-b) and N-terminal lid domains (c-d) were performed in PyMOL. Peptides are shown as cartoons and ADP (from 4AXE) is shown as sticks (pink, coloured by element).

The clasp which encloses the active site when InsP ligand is bound, forms between contacts between $\alpha 6$ (Leu128-Ala131) and L3 (Gly251-Ser261) (González *et al.*, 2010). Interestingly, in the WT apo structure the residues E18, R130 and W129 interact in a stacking formation, with the W129 also stabilising the 252-256 loop (Figures 3.7a; Appendix 5). These stacking interactions are not possible in the W129A mutant apo structure due to the absence of the critical W129 residue, hence the more open N-lobe lid formation. In the WT ADP structure, this interaction is further strengthened by W129 forming an interaction specifically with E255 (Figure 3.7b; Appendix 5), a contact proposed to only form when an inositol substrate binds (Baños-Sanz *et al.*, 2012a). Comparably, in the equivalent W129A AMP-PNP mutant (PDB 4AXD) or the W129A ADP (PDB 4AXE), the E255 does not form this interaction due to the absence of the tryptophan residue. Unfortunately, no meaningful comparison can be drawn

from the ADP-bound structure (PDB 3UDS), a low-resolution structure where the E255 loop is incomplete despite the presence of electron density in this region. The 3UDS structure is also in a different spacegroup (P1) and therefore, has different cell symmetry contacts, which may affect the positioning of the N-lobe lid.

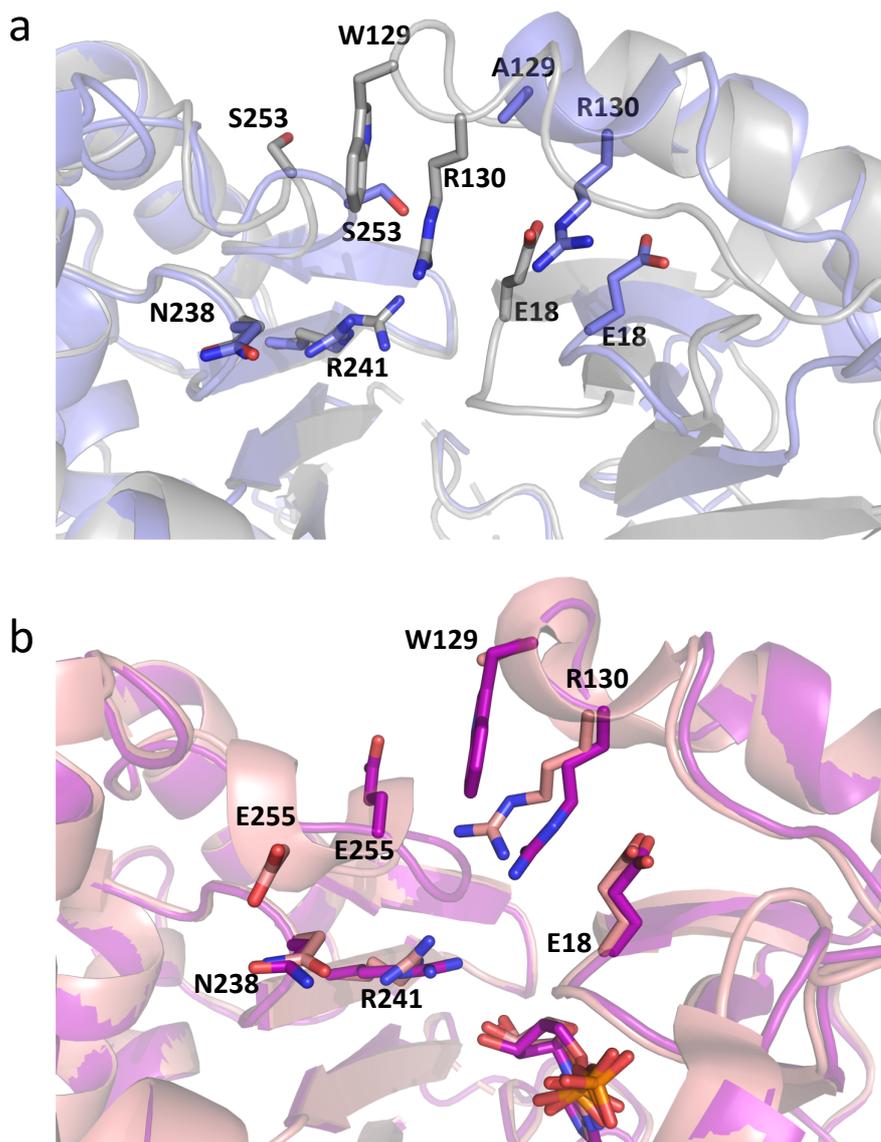


Figure 3.7 Alignments of alike wildtype and W129A mutant structures highlighting clasp forming residues as sticks; where (a) shows apo models, wildtype in silver and 4AXC in slate; and (b) shows ADP bound structures where wildtype is purple and 4AXE is pink.

3.3.4 Investigating the AtIPK1 GRP motif

The Glycine-336/Arginine-337/Proline-338 tripeptide, adjacent to the zinc co-ordinating residues of AtIPK1, was targeted as a potential DNA binding motif. Three GRP SDM mutants, referred to as GKP, SGP and Del(GRP), were chosen to explore the importance of GRP for DNA binding (Figure 3.8). The intended mutagenesis was modelled with Swiss model in order to predict potential structural differences, these show that minor changes may not impact structural integrity, despite Del(GRP) mutant truncation, zinc co-ordinating residues are predicted to remain less than or equal to 2.5 Å contact distance of zinc extracted from aligned wildtype structure.

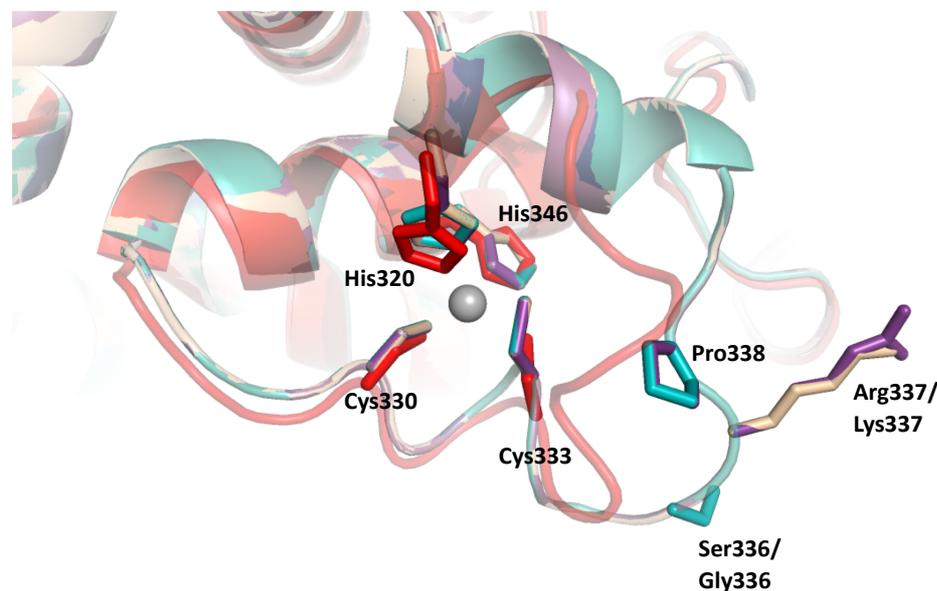


Figure 3.8 Swiss models of wildtype AtIPK1 (purple) and SDM mutants: GKP (taupe), SGP (teal) and Del(GRP) (red) represented as cartoon and aligned in PyMOL. Expanded view of GRP motif of wt and mutant kinases, with RMSD of 0.000 between wt and gkp and sgp mutants; and 0.023 RMSD between wt and del(GRP). Residues of interest shown as sticks.

3.3.4.1 Production of active AtIPK1 GRP mutants

SDM was achieved using *pEAQ-HT-GFP[CaMV 35S-AtIPK1]* with mutagenic primer pairs, and gradient PCR was used to find the optimum annealing temperature (50 °C) for each construct. Verification of each mutation was confirmed by amplification of the insert (~1400 bp) with AtIPK1 primers (Figure

3.9a) and gel electrophoresis, which was also used to approximate the ratios of double digested pOPIN F (cut with *Kpn*I and *Hind*III) and insert fragments, for In-fusion cloning. Successful clones were identified with blue-white screening, where white colonies were picked, and confirmed by DNA gel electrophoresis (Figure 3.9b) and sequencing with pOPINF primers (Appendix 6). SDM constructs were finally transformed into *E. coli* expression strains (BL21 (DE3), Rosetta2 (DE3), SHuffle T7 Express and SHuffle T7 LysY).

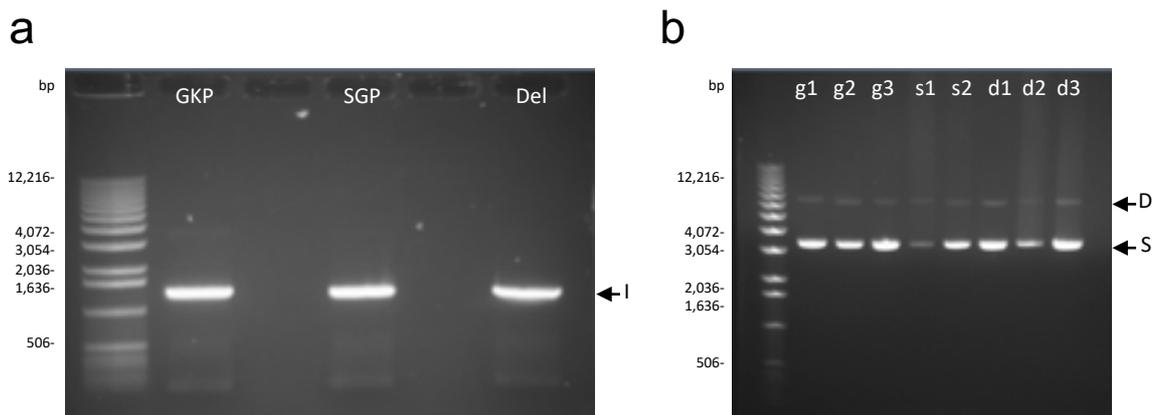


Figure 3.9 Verification of (a) SDM AtIPK1 fragments amplified with pOPINF primers for In-fusion reaction, and (b) plasmid preparations of SDM pOPINF constructs. Where black arrows show: I = AtIPK1 insert (expected mass: ~1400 bp), D = damaged plasmid DNA, including nicked or relaxed or linear DNA (expected mass of ~6800 bp) and S = supercoiled plasmid DNA moves more quickly through agarose gel, is expected to appear smaller and brighter on the gel. AtIPK1 mutants are labelled as follows: g = GKP, s = SGP and d = Del(GRP) mutant, the numbers correspond to different preparations.

Expression trials were conducted to find the most suitable strain for production of AtIPK1 SDM mutants (Figure 3.10; Table 3.4). All SDM mutants expressed well in SHuffle® T7 LysY cells, whereas in other strains the deletion mutant expressed poorly or protein was insoluble. Therefore, this strain was chosen in scaled up 6 L expression cultures for protein production of AtIPK1 mutants to produce enough kinase for characterization and comparisons with wildtype AtIPK1. The proteins were purified using the same steps as used for wildtype AtIPK1, including removal of nucleotide with heparin chromatography (Figure 3.11; Appendix 7); however, protein yields were lower than expected for GKP and

Del(GRP) mutants (200 μ L of 4 mg/mL and 100 μ L of 4 mg/mL respectively) providing challenges for *in vitro* assays.

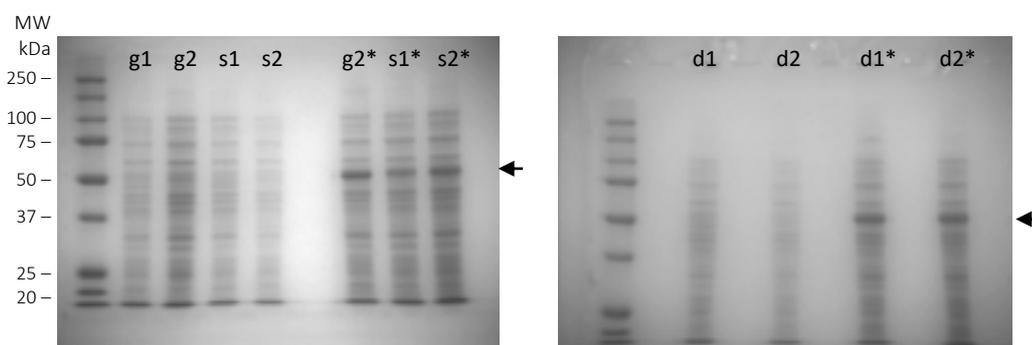


Figure 3.10 SDS-PAGE results of representative expression trial for AtIPK1 SDM mutants in *E. coli* SHuffle T7 LysY. Samples were run on 12% polyacrylamide gels and total protein was visualized with InstantBlue™, a black arrow indicates AtIPK1. Lanes labelled with ‘g’ are GKP mutant, ‘s’ are SGP mutant and ‘d’ are Del(GRP) mutant, and * indicates overnight induction with 0.2 mM (odd numbers) or 0.4 mM IPTG (even numbers).

Table 3-4 Summary of expression data for pOPINF[T7lacO:AtIPK1] GRP mutants ON induction in different *E. coli* strains.

Construct	Strain	[IPTG] (mM)	Expression	Solubility
GKP	BL21 (DE3)	0.2	++	N/A
	Rosetta 2 (DE3)	0.2	+	-
	SHuffle T7	0.2	++	N/A
	Express			
	SHuffle T7 LysY	0.2-0.4	++	N/A
SGP	BL21 (DE3)	0.2	++	N/A
	Rosetta 2 (DE3)	N/A	N/A	N/A
	SHuffle T7	0.2	+	N/A
	Express			
	SHuffle T7 LysY	0.2-0.4	++	N/A
Del(GRP)	BL21 (DE3)	0.2	-	-
	Rosetta 2 (DE3)	N/A	N/A	N/A
	SHuffle T7	0.2	++	+
	Express			
	SHuffle T7 LysY	0.2-0.4	++	N/A

- is no expression/solubility

+ is expression/solubility detectable

++ is high expression/highly soluble protein.

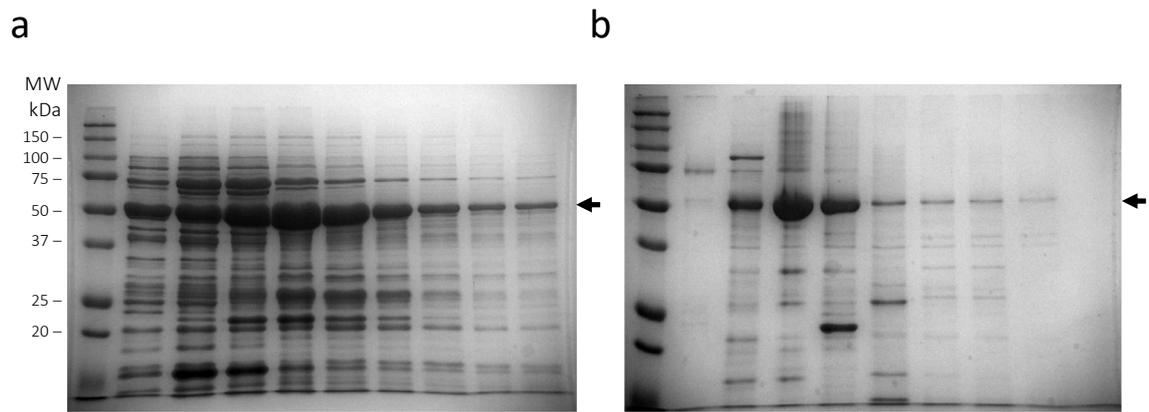


Figure 3.11 Example SDS-PAGE analysis of AtIPK1 GKP purification. By sampling elution fractions collected from (a) Nickel column and (b) Heparin column. Arrow indicating ~50 kDa AtIPK1.

3.3.4.2 AtIPK1 GRP mutant kinase activity

To compare the DNA binding activity of different kinases, it should first be shown that enzymes are uniform in other qualities. Having yielded sufficient amounts of SDM AtIPK1 for analysis, the mutants were assessed alongside wildtype AtIPK1 for purity, kinase activity and substrate binding. Firstly, the purity and quantity of protein was assessed using SDS-PAGE analysis on a 12% polyacrylamide gel (Figure 3.12). The 50 kDa protein was shown to have similar quantities amongst wildtype and mutants to be conducive to comparisons, variations were likely due to human error in loading because there is no consistent pattern of differing protein presence between the two concentrations used. The gel was overloaded with AtIPK1 to aid detection of impurities, most clear contaminant is the appearance of a lower molecular weight band within the Del(GRP) 5 μ M sample. This contaminant is also present in lesser amounts in SGP and wildtype samples.

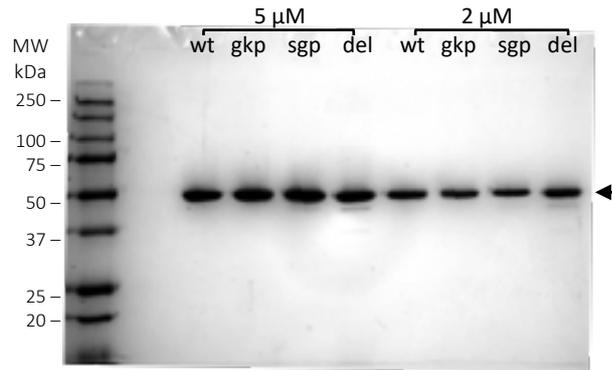


Figure 3.12 SDS-PAGE analysis of purified AtIPK1 SDM mutants alongside wildtype (wt) on a 12% polyacrylamide gel; a black arrow shows AtIPK1 (~50 kDa).

The kinase activity of AtIPK1 GRP mutants was monitored using HPLC following the conversion of ATP to ADP at A258 (Figure 3.13), because the stoichiometry of substrate (InsP₅ and ATP) to product (InsP₆ and ADP) conversion by AtIPK1 is a 1:1 (Sweetman *et al.*, 2006). The initial rate of the reaction can be determined when product formation is at < 10%, and so different concentrations of AtIPK1 were trialled to obtain the reaction rate before this limit was reached. Using 250 nM of protein, the rates for activity of wildtype and SDM mutants are up to 15% conversion, and at 500 nM up to 25% conversion. However, at the lower concentration of protein, there was greater error observed between replicate runs likely due to pipetting error of small volumes from protein stocks. The 95% CI of reaction rates overlap between wt (0.74 to 1.12 %/min) and mutant proteins reaction rates, which means that they are not statistically different and therefore we may consider that the AtIPK1 GRP mutants are functional enzymes.

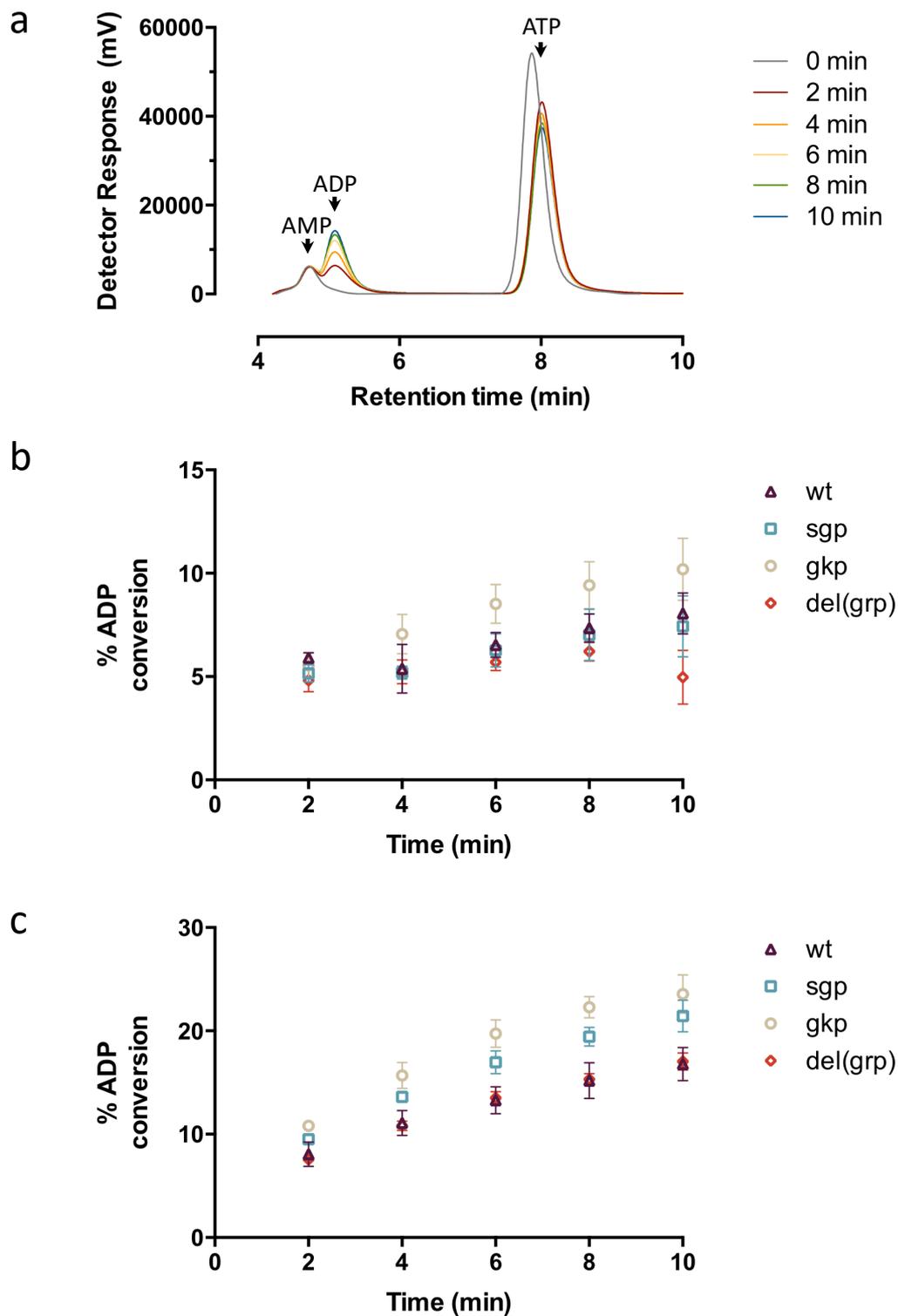


Figure 3.13 Activity of AtIPK1 SDM mutants monitored by ATP to ADP conversion using HPLC. Where (a) shows an example of HPLC trace data for 500 nM AtIPK1 GKP mutant, following ADP production, the mean response of triplicate repeats for each time point as a single trace; (b-c) shows the mean ATP to ADP conversion \pm SD of 3 replicate runs using (b) 250 nM, or (c) 500 nM protein. The different proteins are shown as follows, wildtype (purple, open triangles) and SDM mutants: GKP (taupe, open circles), SGP (teal, open squares), and Del(GRP) (red, open diamonds).

Lastly, InsP-substrate binding of wildtype and SDM mutants were monitored by FP (Figure 3.14), using the substrate analogue 2-FAM-InsP₅ (Appendix 1) in binding buffer containing 100 mM NaCl, to limit non-specific interactions. At first glance, the curves of the saturation are very similar for all proteins. The binding affinities calculated using a variable slope model in Prism v6.0 (GraphPad) show that there is approx. 3-5% difference between EC₅₀ values of wildtype, GKP and Del(GRP) mutants. This is likely due to human error and not innate differences in substrate binding. The outlier in this, is the SGP mutant, which has approx. 14% higher affinity for 2-FAM-InsP₅ than wildtype protein, which is a surprising result as the target tripeptide is not close to the active site/presumably where 2-FAM-InsP₅ binds. With attention to the points of the generated curves, there are not many protein concentrations plotted between 0 to 100% saturation, which may increase the margin of error between EC₅₀ calculations. This may account for the greater difference observed by the SGP mutant or suggest that the error limit for this experiment is at least 14%.

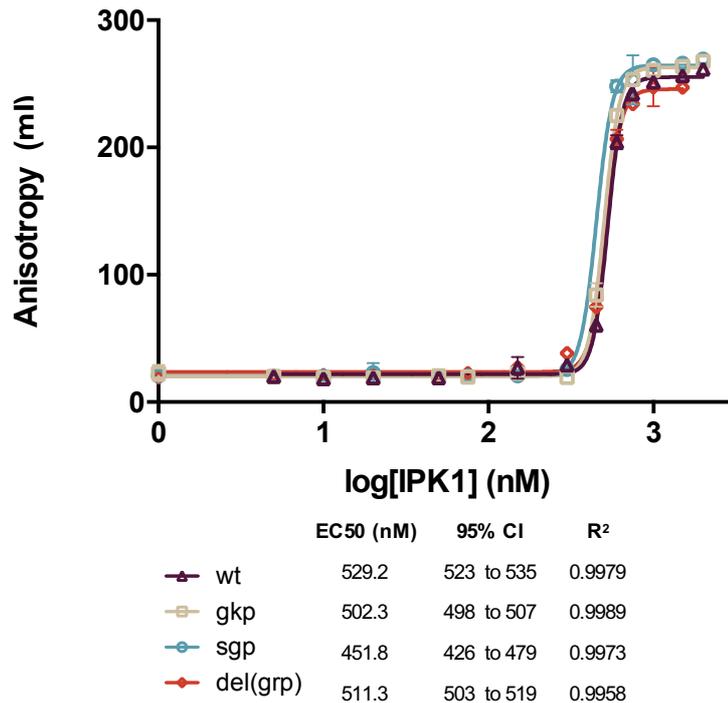


Figure 3.14 Wildtype AtIPK1 and SDM mutants bind to 2-FAM-InsP₅ (2 nM), monitored by FP assay. Each point is the mean \pm SD of ≥ 3 replicates, and FP data was processed using Prism v6.0 (GraphPad) with a variable slope model. Each binding assay was performed with 100 mM NaCl, and incubated for 1 h before plate readings at 25 °C. The different proteins are shown as follows, wildtype (purple, open triangles) and SDM mutants: GKP (taupe, open squares), SGP (teal, open circles), and Del(GRP) (red, open diamonds).

3.3.5 AtIPK1 GRP mutants binding DNA

Having established that the enzyme activity and inositol binding affinities of the wildtype and mutants were not greatly impaired, comparisons of DNA binding were conducted with the limited amount of protein available using FP (Figure 3.15) and F-EMSA (Figure 3.16). Both assays showed that GRP motif mutants bind DNA. Firstly, FP experiments were performed with FAM-12mer in a binding buffer with 100 mM or 50 mM NaCl. In the former condition, complete binding of the nucleotide oligomer was not observed (Figure 3.15a) and binding affinities were unattainable. To overcome this, three discreet concentrations of protein were used as points of comparison to assess binding (Figure 3.15b). The FP values obtained show little differences between the wildtype and SDM mutants (shown

by overlapping error bars, representing SD), although at every tested protein concentration the wildtype has the highest FP values. This may be due to errors in pipetting rather than actual differences. To exaggerate differences by making the curves more gradual and to allow full saturation to be achieved, a final FP assay was performed with 50 mM NaCl in the binding buffer (Figure 3.15c). Unfortunately, the GKP mutant was not included in this experiment because of limited supply of the protein. The wildtype and SGP curves are identical in this final binding assay, with EC50 values (233.4 nM and 243.5 nM respectively) changing less than 5%. The Del(GRP) mutant performs similarly, yet its curve is slightly shifted to the right and an EC50 value of 324.9 nM (95% CI, 313 to 338 nM) was obtained, an approx. 1.3x lower binding affinity. This difference could be due to pipetting error, or perhaps shows that the Del(GRP) mutant was more degraded than either wildtype or SGP protein, although such a difference is unexpected when considering there was only a 5% observed difference in 2-FAM-InsP₅ binding affinity (Figure 3.16). There is also the possibility that the Del(GRP) mutant has a lower binding affinity for the DNA; despite this capacity clearly not being abolished. In any case, the GRP motif of AtIPK1 does not appear to be essential to DNA binding.

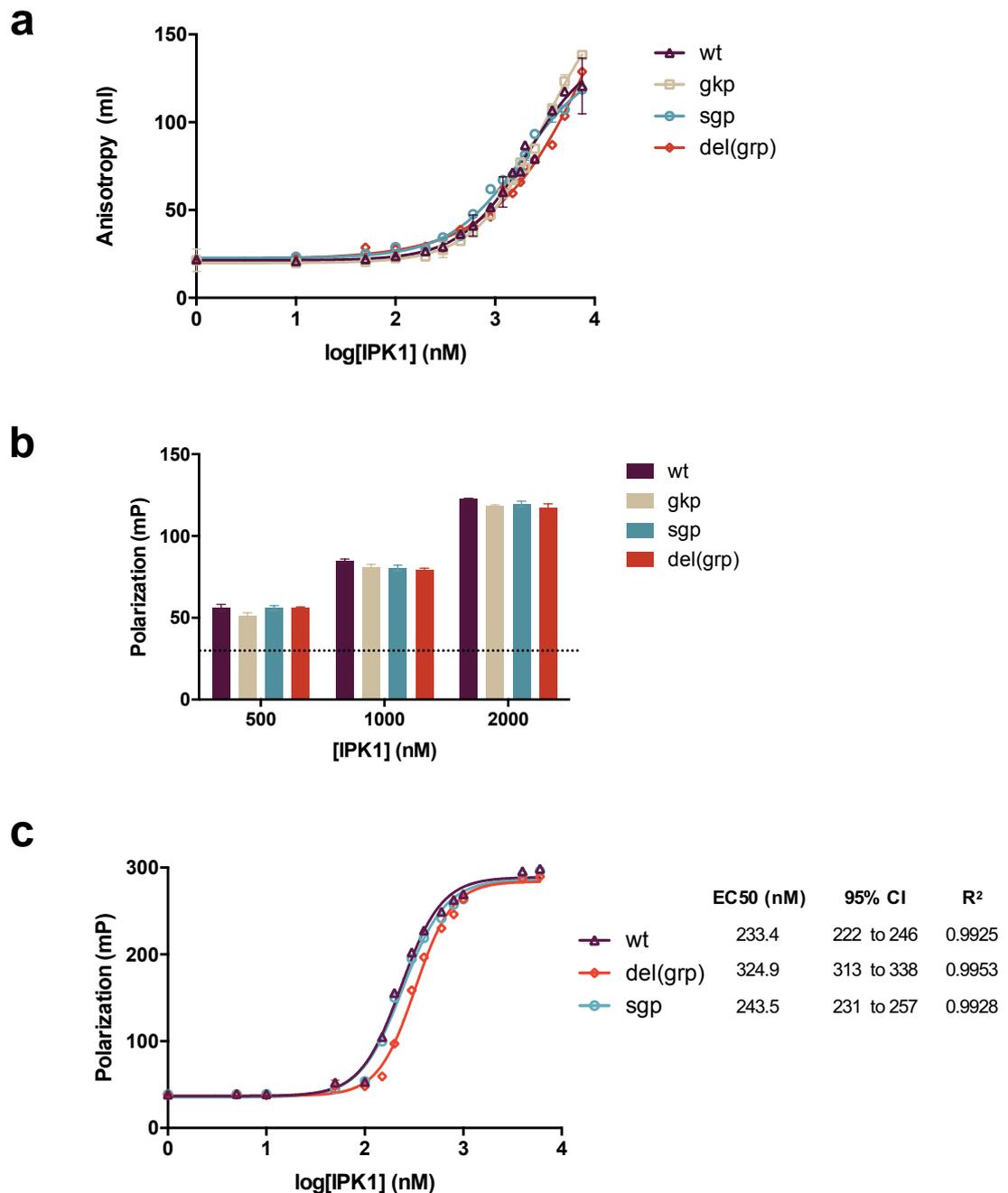


Figure 3.15 Wildtype AtIPK1 and SDM mutants bind to 2 nM 5'-FAM-12mer in FP assays. Presented as: (a) a variable slope model of AtIPK1 with DNA and 100 mM NaCl in binding buffer; (b) FP at discrete concentrations of AtIPK1 with DNA with 100 mM NaCl in binding buffer, where the dotted line is the average FP when of probe alone; and (c) a variable slope model of AtIPK1 with DNA and 50 mM NaCl in binding buffer. Each point is the mean \pm SD of ≥ 3 replicates, and the data were processed using Prism v6.0 (Graphpad). In each assay, binding reactions were incubated for 1 h before plate readings at 25 °C. The different proteins are shown as follows, wildtype (purple, open triangles) and SDM mutants: GKP (taupe, open squares), SGP (teal, open circles), and Del(GRP) (red, open diamonds).

AtIPK1-DNA binding visualized by F-EMSA were likewise very similar when observing wildtype and mutants (Figure 3.16). The same DNA probe was used with discrete concentrations of protein to determine the amount of fluorescent DNA bound to the protein (Figure 3.16a). It is clear in the first instance that although all proteins bind to the FAM-DNA, a large proportion of DNA is not bound in the presence of 1 or 5 μM protein. To assess the amounts of protein loaded onto the gel, a Coomassie stain was used after the first visualization of the gel (Figure 3.16c). When these gels are overlaid (Figure 3.16b), the bands can be clearly identified and little difference is seen between the wildtype AtIPK1 and the mutants. At 5 μM of protein it is clear however that the wildtype mutant has been underloaded compared to the fairly equal amount of SDM mutants in neighbouring lanes. This proportionally skews the amount of wildtype complex observed at that concentration. The Del(GRP) mutant also has a comparatively low amount of DNA complex, despite not being underloaded. This may suggest lesser binding affinity for Del(GRP), however further experiments should be used to confirm this.

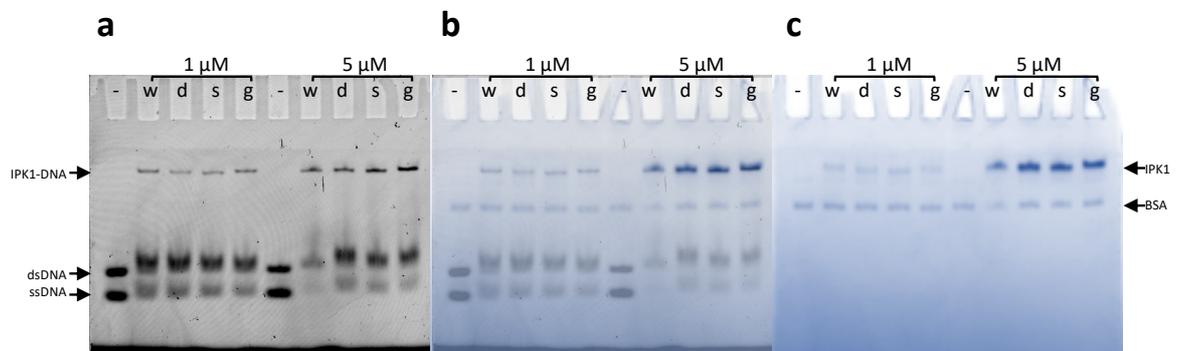


Figure 3.16 F-EMSA analysis with wildtype vs SDM mutants of AtIPK1 using FAM-12mer on an 8% NATIVE gel. Components of the binding reaction were visualized as follows: firstly (a) the DNA by FAM detection, (c) protein detection with InstantBlue™ and finally (b) the two images were combined with 0.7 opacity. Notation of 'w' for wildtype AtIPK1, 'd' for del(grp), 's' for sgp and 'g' for gkp mutants is used.

3.3.6 Conformational changes of AtIPK1 and ligand binding

Previously in the Brearley lab, high-throughput screening using FP was performed to identify inhibitors for moderating IPK kinase activity (Whitfield *et al.*, 2018). In a similar manner, FP displacement was utilized with different AtIPK1

ligands as potential inhibitors of the AtIPK1-DNA interaction (Figure 3.17). Titration of physiological substrates and products against AtIPK1-bound FAM-DNA yielded different results: nucleotides (ATP and ADP) did not interfere with DNA binding, whereas InsPs (InsP₅-2'OH and InsP₆) displaced the probe. It appears that InsP₅-2'OH (IC₅₀ 454.4 nM; 95% CI, 433.7 to 474 nM) was the more efficient inhibitor, contrasting with the steeper curve generated by addition of the InsP₆ (IC₅₀ not calculated by model) when DNA binding is affected but only after addition of >100 μM inositol phosphate. The active site inhibitor purpurogallin was also an effective inhibitor of AtIPK1-DNA binding (IC₅₀ not calculated by model), however, the DNA minor groove binder netropsin did not displace the probe.

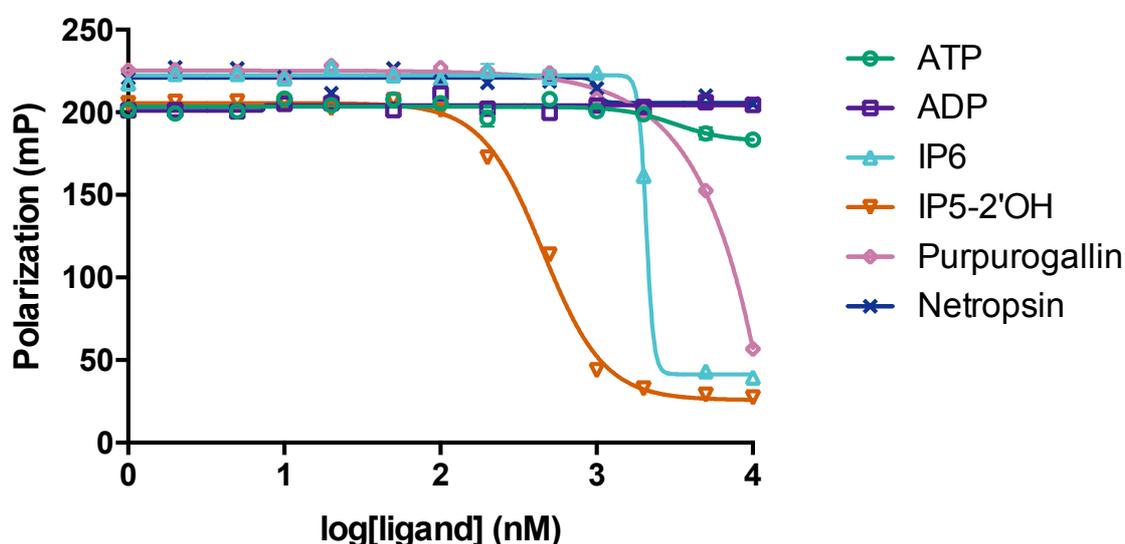


Figure 3.17 Displacement of the AtIPK1-DNA complex by different ligands using FP with 500 nM protein and 2 nM DNA. Each point is the mean \pm SD of ≥ 3 replicates, and the data was processed using Prism v6.0 (Graphpad) with a 4-parameter fit. Each binding assay was performed with 50 mM NaCl and incubated for 1 h, before plate readings at 25 °C. Ligands are presented as follows: ATP as green, open circles; ADP as purple, open squares; InsP₆ as teal, open triangles; InsP₅-2'OH as red, open upside-down triangles; purpurogallin as pink, open diamonds; and netropsin as navy crosses.

3.3.7 Investigating AtIPK1-DNA binding *in silico*

In the absence of diffraction data of the protein-DNA complex, DNA-binding can hypothetically be modelled. Firstly, the electrostatic potential across the surface of the protein was inspected using the PyMOL APBS software (Appendix 8) and

the non-homology dependent webserver BINDup (Paz *et al.*, 2016) was run to predict a DNA binding site by identifying the largest positively charged binding interfaces of AtIPK1 using the physicochemical properties of adjacent surface residues (Figure 3.18a). Unsurprisingly, a large, concentrated pocket of positively charged residues, including lysines, dominates the AtIPK1 active site which may expand into a larger binding site for protein or nucleotide interactions. The structure of the non-DNA binding protein, AtITPK4 was similarly inspected, wherein the InsP pocket was also highlighted as a potential positive binding surface, however, was far less expansive (Figure 3.18b). In both cases, the IPKs were not predicted to be nucleic acid binding proteins. Furthermore, the DISPLAR server (Tjong and Zhou, 2007) was used to predict DNA binding residues, in a similar approach of finding adjacent residues where there is enrichment of positively charged residues however also considering sequence conservation as a feature of interface residues. There was overlap between putative binding surface and putative DNA interface residues, particularly around the InsP co-ordinating residues (Figure 3.18c). The positive InsP binding pocket and co-ordinating residues are unsurprisingly highly conserved amongst plant IPKs, as illustrated using Consurf-DB (Chorin *et al.*, 2020) [Appendix 9], therefore this is not necessarily a feature for DNA binding. Excluding this site, smaller clusters of positively and negatively charged surfaces are seen across the structure, however there were no predicted binding sites close to the GRP tripeptide.

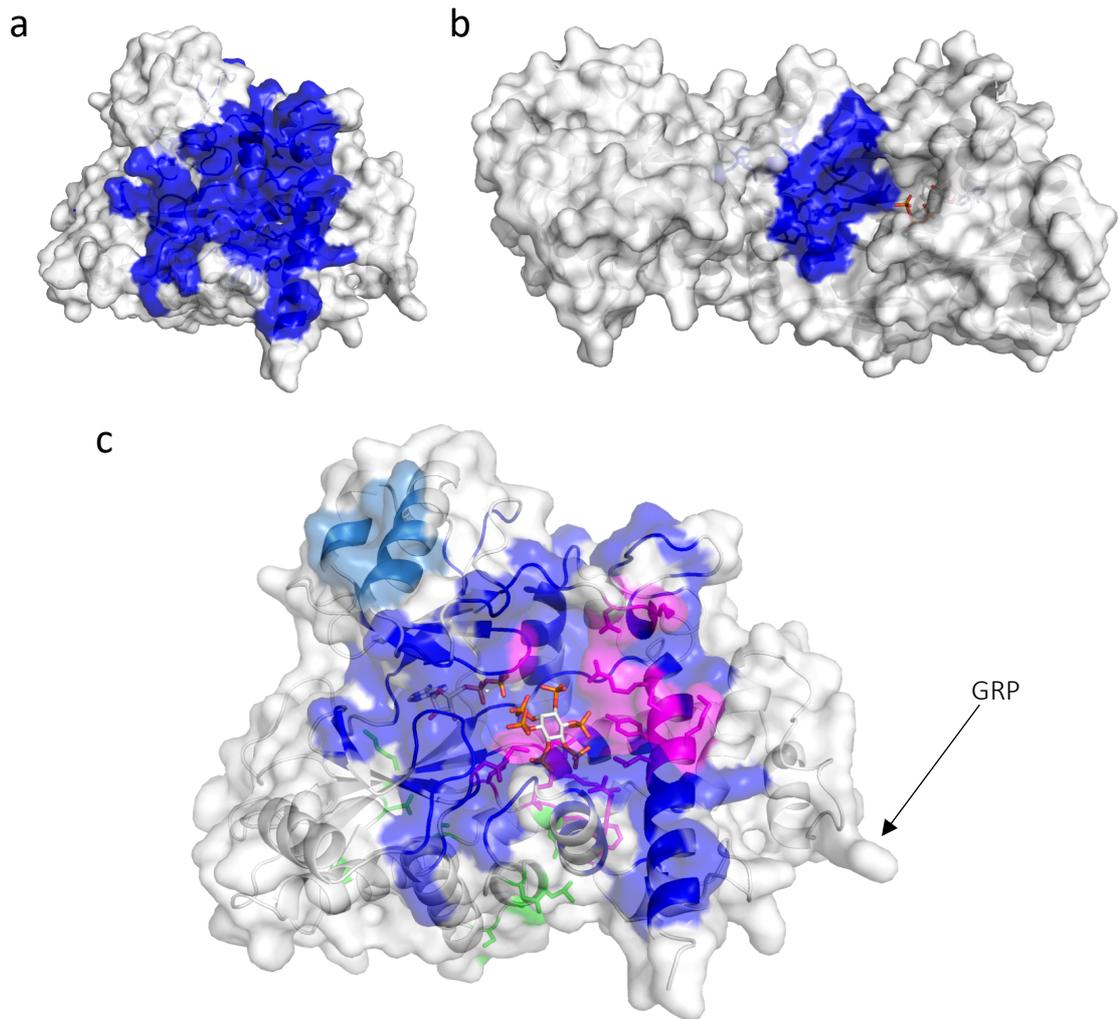


Figure 3.18 Predicted positive binding surface (blue) of (a) AtIPK1 and (b) AtITPK4 from BINDup, with (c) additional features annotated for AtIPK1; including: a second positive patch (sky blue) and DNA binding residues predicted from DISPLAR (sticks, green and magenta where overlap occurs between BINDup analysis).

With particular focus on the putative DNA binding motif, an alignment was made with the GRP motif from AtIPK1 and an AT hook peptide from a peptide-DNA complex (PDB 3UXW), which uses the same dodecamer (12mer) as synthesised in this study (Figure 3.19). In this alignment, R337 appears to contact DNA in the minor groove, however major clashes are present in the modelled conformation. There appears to be an adjacent hook, residues Gly181-Lys182-Glu183, which may contact the major groove of DNA. Apart from the highly conserved zinc co-ordinating residues (H320, C330, C333 and H346) there is low evolutionary conservation around the GRP loop (resi 326-345) and GKE (resi 182-

183) (Appendix 9) which may point to a lack of functional importance of these residues (Luscombe and Thornton, 2002).

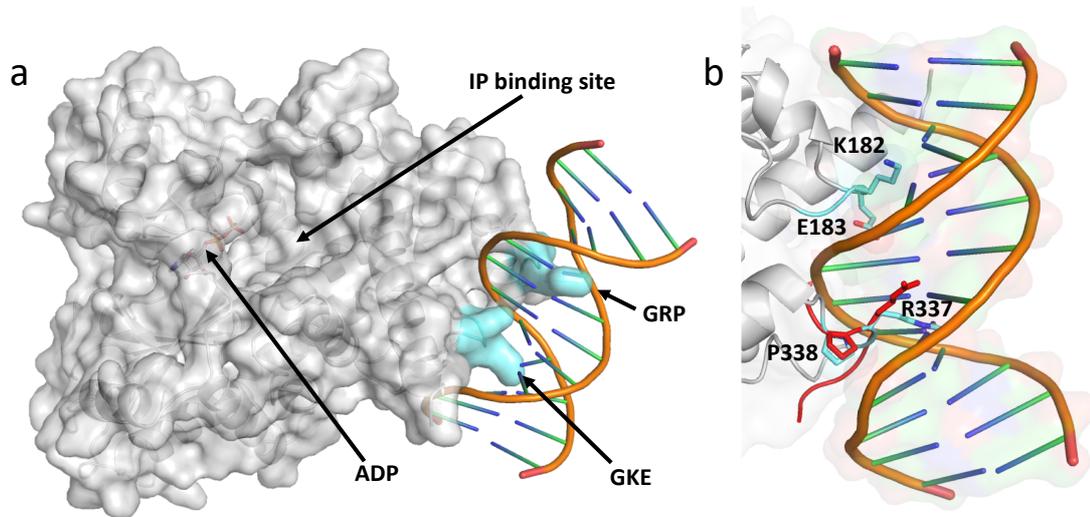


Figure 3.19 Alignment of GRP hook peptide (in complex with duplex dodecamer, PDB 3UXW) and WT AtIPKI-ADP.

To reduce the bias inherent from targeting a putative motif without experimental evidence, HDOCK (Yan *et al.*, 2017) was used to generate AtIPKI-DNA docking models using wildtype apo and ADP bound AtIPKI structures, the AtIPKI-substrates ternary complex from high resolution models (PDB 2XAM and 6FJK), and ds dodecamer [12mer] (from PDB 3UXW). This server produces 100 models of protein-DNA complexes from template-based modelling and/or *ab initio* methods to identify potential DNA binding residues (<5 Å distance apart). From this search, the top 10 models (with docking scores ranging -230.34 to -174.18) were inspected, showing multiple potential binding sites on the AtIPKI surface (Appendix 10). For each AtIPKI template, the highest scoring docking models (-230 to -207) predicted DNA binding across the active site (Figure 3.20a-b); including the involvement of InsP co-ordinating residues such as R45 and R130. Amongst the top models of AtIPKI-DNA complex, 25 out of 40 models showed DNA binding across the active site and 10 models showed DNA co-ordinating residues within the zinc binding region (residues 320-346), two of which (from using the apo AtIPKI template) where R337 contacts DNA within the minor groove (Figure 3.20c-d). Separate from these typical IPKI distinctive features, five models demonstrate DNA binding including approx. residues 376-404 sharing

overlap with predictions from the DISPLAR server. In all cases from HDock modelling, many binding residues have been identified (≥ 17 residues) which may be involved in interacting with the DNA sugar phosphate backbone as well as potentially specific interactions with DNA bases.

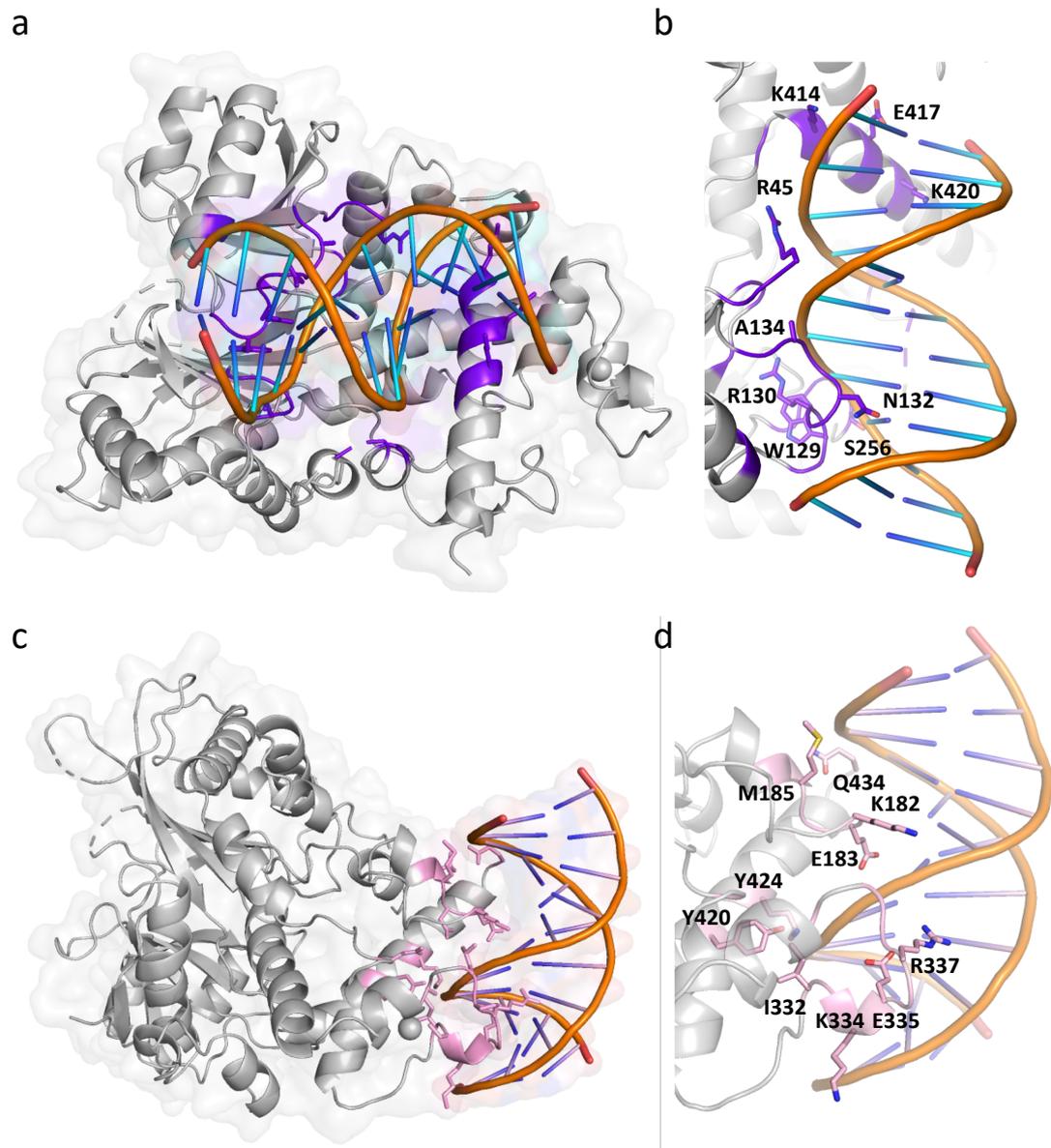


Figure 3.20 Two disparate models of DNA docking to wildtype apo AtIPK1 generated from the HDock webserver. Firstly showing (a,c) the overall docking of DNA (PDB 3UXW) with surface and cartoon representations of AtIPK1 and DNA, with designated DNA binding residues ($<5 \text{ \AA}$) highlighted in (a,b) model 1 (docking score = -210.16; purple) with binding across the active site, and (c,d) model 9 (docking score = -174.74; pink) with co-ordinating residues including the GRP tripeptide. Secondly (b,d), a zoomed in view of DNA binding residues $<3 \text{ \AA}$ are highlighted in each model as sticks.

In the previous chapter, AtIPK1 was shown to form a dimer in solution and appeared to bind DNA in either monomer or dimer oligomeric state (section 2.3.4). Here, the webserver GalaxyHomomer (Baek *et al.*, 2017) was used to predict the dimer interface from the monomeric AtIPK1 structure and produced five models, showing possible orientations/surfaces of interaction from the structural template and *ab initio* docking. Amongst these models, the only template-based ‘model 1’ using PDB 4AXD (SEQ ID: 97.2, TM-score: 0.9863) was highly similar to the symmetry mates generated from apo or ADP-bound AtIPK1 models in PyMOL (Figure 3.21), where all other models were generated *ab initio* and had docking scores between 1131.135 and 908.076. In such a dimer conformation, both GRP motif and substrate binding site appear to be accessible. The HDOCK server (Yan *et al.*, 2017) also has a homomer prediction feature, whilst requiring symmetry group as an input (i.e. C2 for a homodimer), where 100 models are generated and scores calculated for the top ten (docking scores: -257.70 to -184.06). From these results, ‘model 6’ was highly similar to the ‘model 1’ predicted by GalaxyHomomer (RMSD = 0.640, 858 common atoms) in alignments (Appendix II).



Figure 3.21 Model of AtIPK1 homodimer (pink) from GalaxyHomomer in alignment with wt apo AtIPK1 structure (grey) and symmetry mate (black).

3.4 Discussion

Within this study, the AtIPK1-DNA complex was investigated by attempting X-ray crystallography. Whilst unable to obtain crystals with the protein-DNA complex, two wildtype AtIPK1 structures were solved including a novel apo structure, which upon inspection of the flexible N-lobe, aligned more closely to the nucleotide bound rather than apo W129A mutant structure and challenges the current model of the open conformation AtIPK1 described by González *et al.*, 2012. Without a structure of the AtIPK1-DNA complex, a putative DNA binding motif, the GRP tripeptide, was investigated using SDM to generate GKP, SGP and Del(GRP) AtIPK1 mutants. Subsequently, comparative DNA binding between wildtype and mutant proteins were assessed using fluorescence polarization (FP) and fluorescence-based electrophoretic mobility shift assays (F-EMSA), thus revealing that this tripeptide is not essential to DNA binding activity. FP was further utilized to explore potential inhibitors of AtIPK1 DNA binding, showing that InsP ligands, as well as the active site inhibitor purpurogallin inhibit DNA binding activity. Lastly, *in silico* modelling using available and functioning web servers to provide possible DNA binding residues and potential targets for future work.

3.4.1 Crystallization of protein-DNA complex

Attempted crystallization of the AtIPK1-DNA complex did not produce a structure, most likely due to the difficulties faced by crystallographers working with DNA. At the time of writing, there are 9,715 DNA-protein complex structures available in the Protein Data Bank (PDB), compared to 159,817 of protein and 2,106 of DNA alone structures out of a total 182,9491 entries. Despite annual increases in number of protein-DNA complexes published, there are many reasons that these complexes are only ~5% of entries in the PDB (Berman *et al.*, 2002). Fundamentally, in order to successfully determine macromolecular structures with X-ray crystallography, diffraction-quality crystals must be produced (Hollis, 2007). In contrast to crystallization screens for traditional single species systems for which parameters include continuous variables such as protein concentration, pH of solution and temperature of growth (Russo Krauss

et al., 2013); the crystallization of protein-DNA complexes present a greater challenge due to an additional set of variables which complicate the generation of homogenous samples and crystal formation. For example, the purity and concentration of each individual component, protein:DNA ratios, length and composition of DNA (Hollis, 2007). In fact, choosing the right DNA may be more important than exploring condition diversity (Watts, 1993).

Typically, it is recommended that broad screening is used with different DNAs and different conditions, followed by directed screens and optimization to produce better quality DNA containing crystals (Hollis, 2007). Unfortunately, we did not have access to large scale screen equipment/automated screen monitoring required for such large-scale screens. In this study, the palindromic 'AT-hook' dodecamer (12mer) was synthesized and annealed to be used as a candidate for AtIPK1 binding. This 12mer has previously been used successfully in crystallography trials with an AT-hook peptide motif (Fonfría-Subirós *et al.*, 2012). An annealed palindromic nucleotide sequence allows protein to bind either strand and remain structurally homogenous within a lattice (Hollis, 2007). Shorter length DNA is preferred in crystallization trials because it has smaller number of potential binding positions across the protein and may prevent inhibition of crystal nucleation (Hollis, 2007). In addition, tight binding of the 12mer with AtIPK1 (K_d within micromolar range) has been demonstrated through FP (Section 2.3.1). The stability of the complex is also important to diffraction quality (Hollis, 2007). However, we were not able to demonstrate conditions in which the AtIPK1-DNA complex were homogenous, nor improve the annealing protocol to generate homogeneous dsDNA. In future attempts, dsDNA could be purified before crystallization trials to improve homogeneity.

Another consideration for two-species crystal screening is the stoichiometry of the protein-DNA complex. If there is a ratio of single protein molecule to single dsDNA, it is suggested to start screening with a slight excess of DNA (e.g. 10%), however if there are multiple protein subunits or binding sites on DNA, an excess can be detrimental (Hollis, 2007). In the previous chapter, AtIPK1 binding DNA in

1:1 ratio was observed using low protein concentrations, however the possibility of an AtIPK1 dimer-DNA interaction was not discounted. Since high concentrations of protein are required for crystallography screens, monomer and dimer AtIPK1-DNA screens were used in crystal screening set-up, resulting in high demands of protein and DNA for each experiment. Further difficulties may arise when binding is not sequence specific, as a single position of binding is important for the homogeneity of the crystal lattice (Hollis, 2007), yet the AtIPK1-DNA interaction observed thus far has been non-specific (Section 2.4.2). Future work may involve screening oligonucleotide libraries to find preferential binding sites of AtIPK1 and optimizing component ratios using EMSAs (Hollis, 2007). This would allow a more directed crystal screening approach.

3.4.2 ADP and apo models of wildtype AtIPK1

Despite a lack of AtIPK1-DNA complex structure, two wildtype protein structures were solved: the ADP-bound and a novel apo AtIPK1. The presence of the ADP yielded from initial screening, was an unexpected find. It has not been previously reported that AtIPK1 co-purifies with nucleotide and the addition of a nucleotide stripping step was required to subsequently crystallize the apo protein. The use of a heparin column in AtIPK1 purification is not novel however, as methods have been published (Baños-Sanz *et al.*, 2012b). The ADP found in the active site of the wildtype protein (Figure 3.4) was possibly an artefact from the DNA annealing buffer, which contains ATP (which may degrade over time to ADP as first hydrolysis product), however it is also possible that ADP was pulled from the *E. coli* expression host system. This latter scenario helps account for the apo structure being solved after heparin purification was performed.

Up to this point, the W129A mutant was generated to solve the apo structure of AtIPK1 (PDB 4AXC) due to difficulties in finding conditions to crystallize the protein without ligands (Baños-Sanz *et al.*, 2012b). The W129A apo model was important evidence for the conformational changes proposed in the activation of the kinase, capturing the 'open' conformation and demonstrative of the movement of the N-terminal lid domain opposed to the nucleotide or InsP bound

form of the protein (Baños-Sanz *et al.*, 2012a). However, in this current study an open conformation was not apparent from the wildtype apo structure (Figure 3.6). It appears that the targeted tryptophan of the W129A mutant, which interacts with E255 forming a clasp upon InsP binding to stabilize the closed conformation (Baños-Sanz *et al.*, 2012a), may also stabilize L3 without an InsP substrate as shown in both ADP and apo wildtype structures (Figure 3.7). Therefore, we present the possibility that the fully open form of AtIPK1 is an artefact of mutation of the W129 residue, which prevents the contacts positioning the N-lobe lid correctly. Whilst possibly not accounting for the large conformational changes predicted by tryptophan fluorescence experiments upon subsequent additions of substrate (Baños-Sanz *et al.*, 2012b), the flexibility of the L3 loop and N-term domain particularly around the active site is shown between the wildtype apo and ADP bound structures (Figure 3.5). Additionally, an increased stability of the N-lobe may be inferred in the ADP bound structure as the configuration of the L3 loop allows the stacking of E255 and W129, which may improve stability (Sivasakthi *et al.*, 2013); whereas this specific interaction is not seen in the apo form (Figure 3.7). This would agree with the improved N-term stability observations of nucleotide bound protein vs apo in limited proteolysis experiments (Gosein *et al.*, 2012).

In order to avoid overinterpretation of these structural models, we should consider that each model can only provide a snapshot of protein conformation (Burra *et al.*, 2009); and that whilst comparisons of models can aid interpretation of newly solved structures, there are limits to the conclusions that can be drawn (Arkhipova *et al.*, 2017). There can be a number of reasons to explain the differences observed between the newly solved structures and previously published ones, such as point mutations and the presence of different ligands, which have been explored so far in this study. In addition, the quality of models should be considered. Within this study, we provide two relatively low-resolution models of wildtype protein (2.81 Å for apo and 2.23 Å for ADP bound) of a similar resolution to other published AtIPK1 structures (i.e., in the range of 2.02-3.1 Å). As a basic indicator of quality, 2-2.5 Å models are broadly considered more

informative than 2.5-3 Å models, such that the probability of incorrectly placed side chains within flexible regions in the latter category is high (Arkhipova *et al.*, 2017). Therefore, this may down-weight the credibility of details observed from the wildtype apo model, compared to the AtIPK1-ADP structure (Kufareva and Abagyan, 2012). However, density for the side chains responsible for the stacking residues described above have clear density in both models presented here suggesting they are valid snapshots. Optimization of crystallization conditions would be required to obtain higher quality crystals of apo AtIPK1 to solve higher resolution structures and provide a more reliable model (Arkhipova *et al.*, 2017). Divergence between models above RMSD 0.4 Å should not be ignored as natural variation amongst redundant structures and instead may present different conformational states (Burra *et al.*, 2009). Such variance can be due simply to the conditions of crystallization (Mowbray *et al.* 1999) and as such it is difficult to understand conformations of native proteins on a physiological level.

As an alternative and complementary approach to X-ray diffraction experiments, small angle X-ray scattering [SAXS] (Yang, 2014; Kikhney and Svergun, 2015) may be a useful approach. This method is non-destructive and does not require crystalline analytes and reveals the average structure of all the illuminated particles (typically 1-100 nm resolution range) in the bulk material, which may be liquid or solid. SAXS analyses usually result in lower resolution data but may offer a preliminary insight into how AtIPK1 forms an association with DNA and enable improvements of *in silico* modelling.

3.4.3 Investigation of the AtIPK1 GRP motif

In this study, active AtIPK1 GRP mutants (Figure 3.14) were produced in sufficient quantity for preliminary comparative studies. The introduction of point mutations may affect protein production, such as the AtIPK1 C330A and C333A mutants which were insoluble [and this may have been the case for GRP mutants for which target residues are positioned on the same peptide loop (Figure 3.8)], however, soluble protein was obtained. Despite this, the low yields from large scale protein production meant that larger time frames (than available within the

scope of this project) in order to either increase the scale of production or to improve the production pipeline, would be required for experiment replications or elaborations (e.g., displacement assays). In particular, the production of GKP and Del(GRP) mutants were limiting and the former protein was necessarily omitted from a final FP assay. It may also be worth considering that although heparin chromatography was undertaken to strip nucleotides and produce apo proteins for examination, this additional purification step may not be necessary for comparing mutant AtIPK1-DNA binding in future work, because as shown in this chapter, nucleotide binding does not interfere with DNA interactions (Figure 3.19).

Preliminary FP data monitoring DNA interactions with all three mutants against wt AtIPK1 show very similar binding (Figure 3.15). F-EMSA gel images support this further (Figure 3.16), whilst demonstrating that quantification of bands by densitometry is required to prevent human error obstructing observation of subtle differences. Specifically, these data show that point mutation of R337 with K337 does not impact DNA binding, nor does replacement of R337 with G337 (in the SGP mutant) significantly hinder the AtIPK1-DNA interaction. Likewise, the complete removal of the GRP motif did not greatly reduce DNA binding, although in both FP assays and F-EMSA potential lower DNA binding affinity was observed for Del(GRP). Therefore, we conclude that the GRP tripeptide is non-essential to DNA binding affinity and suggest that AtIPK1 mode of DNA binding is dissimilar to AT hook proteins. This is further supported when considering (1) that additions of the minor groove binding ligand netropsin in FP assays did not impair DNA binding (Figure 3.19), whereas AT hook proteins typically contain one or more 'hooks' which bind to the minor groove of DNA with low affinity (Aravind and Landsman, 1998), and (2) the GRP tripeptide is not conserved amongst plant IPKs (Appendix 9). A lack of evolutionary conservation of these residues may suggest low importance to protein function (Luscombe and Thornton, 2002) and specifically points to these residues not being part of backbone-contacting residues (Luscombe and Thornton, 2002), although DNA

binding activity amongst plant IPKIs has not been reported to validate this proposal.

However, the possibility that AtIPK1 residues Gly336-Arg337-Pro338 are involved in DNA binding cannot be ruled out, for example, hits of these residues as putative DNA binding residues were found within top ten models provided from *in silico* docking (Figure 3.18c-d). In these models, a second motif GKE (residues Gly181-Lys182-Glu183) may facilitate DNA binding in the major groove, which may explain why targeting the GRP alone does not abolish DNA binding and that co-operative binding from these motifs may determine sequence specific nucleotide interactions. Future work may involve targeting the GKE tripeptide in conjunction with GRP motif using SDM to generate further AtIPK1 mutants and assessing DNA binding activity and change of sequence specificity of single and double mutants. On the other hand, alternative possibilities for DNA binding sites should be explored in future work.

3.4.4 Regulation of AtIPK1-DNA binding activity

Through DNA displacement assays with AtIPK1 substrates, it was shown that InsP binding inhibits the AtIPK1-DNA interaction with nanomolar to micromolar sensitivity (Figure 3.19). That the EC₅₀ for nucleotide displacement by the physiological substrate, InsP₅, approx. 0.6 μ M is in close agreement with the binding constants obtained for InsP₅ binding (0.35 \pm 0.12 μ M for native protein and 0.42 \pm 0.03 μ M for the W129A mutant) by analysis of intrinsic tryptophan fluorescence (Baños-Sanz *et al.*, 2012a) is compelling evidence that the structural motions that accompany InsP binding are physiologically relevant. The non-charged, purpurogallin, an active site inhibitor that has been shown to induce a closed conformation of AtIPK1 upon binding (Whitfield *et al.*, 2018), also inhibits DNA binding with lower affinity than InsPs. These InsP-induced conformational changes may prevent DNA binding (Tzeng and Kalodimos, 2012). For example, many DNA binding proteins undergo conformational changes upon DNA binding (Mizuguchi and Ahmad, 2014), a reduced flexibility of the AtIPK1 upon

InsP ligand binding may prevent any further conformational changes necessary for DNA binding.

Alternatively it is possible that InsP (or purpurogallin) binding occludes the DNA binding site (or important DNA binding residues), since the AtIPK1 active site is a lysine rich surface which may attract the negatively charged DNA backbone (Luscombe *et al.*, 2001). Within this study, the latter point is supported by *in silico* predictions, which suggest DNA binding overlap with InsP co-ordinating residues (Figure 3.18a-b). Although predicted AtIPK1-DNA docking models had altered possible binding sites when AtIPK1 in different conformations were provided as template.

Since InsP ligand binding is the characteristic function for IPK1 enzymes, the InsP-induced exclusion of DNA binding activity may be highly relevant to physiological regulation. Transcription factors, which regulate the expression of specific target genes, are often regulated themselves through components of the pathways to which they are responsive. If AtIPK1 is involved in specific DNA binding for gene transcription regulation, targets may be linked to InsP metabolic pathways, and the abundance of specific InsP ligands may provide a signal to regulate AtIPK1 DNA binding. Indeed, other IPKs, such as IPMK, have promiscuous enzymatic activity while also showing protein-protein, including protein-transcription factor interactions that may be independent of kinase activity (Malabanan and Blind, 2016). IPMK was first identified as Arg82 (ARGRIII), a regulator of arginine biosynthesis, and has since been shown to be an adapter protein for the master metabolic regulator mTOR (Malabanan and Blind, 2016). Consequently, a case exists for control of metabolic pathways by IPKS. A possible way to explore this in future work may be to investigate promotor regions of IPK or related genes, and screen with binding assays to find whether AtIPK1 exhibits binding preferences to any of these putative targets, compared to randomized oligonucleotides. This may be a step towards finding a cognate sequence for AtIPK1 DNA binding.

3.4.5 *In silico* DNA docking of AtIPK1

To obtain theoretical models of AtIPK1-DNA binding *in silico*, non-homology related inspection of the electrophysical potential of AtIPK1 residues [BINDup (Paz *et al.*, 2016), DISPLAR (Tjong and Zhou, 2007)], consideration of evolutionarily conserved residues [DISPLAR, Consurf-DB (Chorin *et al.*, 2020)] and the hybrid template-based and *ab initio* methods for protein-DNA docking [HDOCK (Yan *et al.*, 2017)] were used. Correspondingly, different putative DNA binding sites were highlighted from these searches (Figure 3.18 and 3.20). One feature identified by each of these methods is the InsP ligand binding site, including all the number one hits from HDOCK modelling (and InsP co-ordinating residues K168, K170, R192, N238, D368, R415, Y419 and K422 from DISPLAR). As the most distinctive feature of IPKs, the extensive InsP-binding pocket includes highly conserved residues and provides a positively charged surface where DNA may bind. However, DNA (or RNA) binding of AtIPK1 was not predicted by the BINDup server; if we assume this a false negative result (as DNA binding has been observed *in vitro*), such predictions may occur when the mechanism of DNA binding does not rely on large positive electrostatic patches or if multimeric complexes of protein are required for DNA binding (Paz *et al.*, 2016). There are a number of AtIPK1 active site SDM mutants, which have previously been generated to verify the importance for ligand binding and enzyme activation (Gosein and Miller, 2013) perhaps DNA binding activity could be explored from these same mutants to help answer whether this feature is important to DNA binding.

Alternative DNA binding sites were modelled by the HDOCK server, where the top ten ranked models were inspected manually and broadly categorized for including one or more of the following: residues involved in InsP co-ordination, proximity to the zinc binding region, clasp formation, the G-loop and 'other' (in which binding sites did not involve AtIPK1 key features). This server uses a combined template and template-free docking approach, wherein the former method is only possible when structurally homologous interactions are available from the PDB, which allows for complex structural flexibility and the latter (from

which the models in this study were generated) is much more limited as rigid body fitting for receptor and ligand is used (Yan *et al.*, 2020). As follows, these docking models do not account for large conformational changes (Yan *et al.*, 2020) which may occur upon DNA binding (Mizuguchi and Ahmad, 2014). Three different conformational snapshots of wt AtIPK1 (the apo and ADP bound structures solved in this study, and ternary complexes from PDB 2XAM and 6FJK) were therefore used as search models, despite differences in structural quality and missing residues. These specific conformational changes may be quite separate from any DNA-induced peptide movements; however, it may be interesting that in closed-conformation approx. 80% of top docking models showed DNA binding across the active site; compared to apo and ADP-bound structures where only three and five out of top ten models respectively encompassed this site. It may be that in regions of greater disorder, protein conformational changes are more likely to occur, but these changes are often induced by specific rather than non-specific DNA binding interactions (Sunami and Kono, 2013). On the other hand, predominantly non-specific DNA binding interactions can have greater impact on the structure of the target DNA; such as bending (Lin *et al.*, 2012) which is specifically difficult to predict in modelling (Yan *et al.*, 2020). These factors remain problems for modelling protein-DNA interactions (Yan *et al.*, 2020), however model improvements can be made by specifying experimental data such as SAXs (Yan *et al.*, 2017).

In this study, the crystal dodecamer (12mer) was chosen as an input for docking models, due to demonstration of AtIPK1 binding this 12mer (see also Chapter 2). However, the structure of this oligonucleotide (from PDB 3UXW) is bent due to specific interactions with the AT-hook peptide. As discussed above, rigid body fitting is commonly used for docking models which exclude possible flexibility of ligand binding. Having demonstrated that the GRP tripeptide is not necessary for DNA binding, it is probable that the observed DNA binding of GRP mutants or wt AtIPK1 occurs through a different mechanism to AT hook proteins. Therefore, it would be interesting to examine different conformations of DNA as input

models for docking models to see whether different potential DNA interacting residues are identified.

Using a similar approach of *in silico* docking, the AtIPK1 homodimer was investigated using the GalaxyHomomer (Baek *et al.*, 2017) webserver and the HDOCK homomer feature, and several models were generated. The structural template-based model from GalaxyHomomer was generated using PDB 4AXD, an AtIPK1[-AMPNP] crystal structure with two monomer subunits (Baños-Sanz *et al.*, 2012a). A highly similar dimer conformation was observed through *ab initio* docking from HDOCK top ten models, where the only restraint of search space was the C2 symmetry input (Yan *et al.*, 2020). Whilst similar limitations (such as lack of experimental supporting data and potentially unknown conformational changes upon dimerization) to model validity apply here, inspection of the template-based dimer reveals that ligand binding may not be obstructed for InsP or DNA (Figure 3.21). From this particular model, the GRP/GKE motifs for DNA binding may align to provide four-point interactions with DNA if the DNA conformation is changed, whereas for DNA binding across the AtIPK1 active site (located on opposite sides of the dimer) would more likely interact with different DNA molecules. In order to validate dimer predictions, mutagenesis may be used for residues that may facilitate dimer formation and dimer formation assessed using AUC analysis, as performed in Chapter 2.

3.5 Conclusion

In this chapter, the mechanism of AtIPK1 DNA binding activity was investigated by X-ray crystallography, DNA binding motif searches, SDM and inhibitor screening. Within the time scale of this project, we were unable to solve a structure of the AtIPK1-DNA complex, and instead provide *in silico* models for potential DNA binding residues. A putative DNA binding motif, the GRP tripeptide, which featured in two of the top ten docking models for apo AtIPK1, was partially characterized by generating and examining three SDM mutants. It was found that this tripeptide is not essential to DNA binding activity, however this does not discount that the GRP loop which comprises part of the ZF motif

may interact with DNA. Inhibitor screening with select ligands, suggest that AtIPK1 does not necessarily act as a minor groove binding protein and that InsP ligands may negatively regulate DNA binding. Future work may involve obtaining experimental data to support any *in silico* models and hopefully characterize AtIPK1-DNA binding by screening oligonucleotide libraries for preferred nucleotide sequences and continued X-ray crystallization screening for the AtIPK1-DNA complex structure, or SAXS analysis to refine docking models, and validation using SDM.

The binding of nucleic acid and displacement by InsPs is an unprecedented result for proteins of this class. While there are a number of examples for the physical interaction of InsPs with nucleic acid interacting proteins for example ADAR2 and ADAT1 (Macbeth *et al.*, 2005) and other proteins that regulate transcription, IPMK's (Malabanan and Blind, 2016) direct modulation of DNA-binding to an IPK by InsPs has not been described. Evidence presented here suggests a mechanism by which the inositol phosphate products of enzymes involved in InsP₆ synthesis can modulate the protein-nucleic acid interactions of IPK1, potentially forming a feedback loop.

4 SPX domain proteins in Arabidopsis

4.1 Introduction

Phosphorus is an essential plant macronutrient, second only to nitrogen in its limitation of plant growth (Goswami *et al.*, 2016). Plant acquisition of phosphorus is severely limited by its availability in the environment (López-Arredondo *et al.*, 2014). Soluble inorganic phosphate (Pi) is the major form by which plants acquire phosphorus, however Pi commonly forms insoluble complexes with other compounds in the soil (Richardson and Simpson, 2011). Plants have evolved several mechanisms to compensate for low Pi availability and maintain Pi homeostasis, these adaptations encompass the phosphate starvation responses (PSR) (Raghothama, 1999). Among a diverse group of proteins involved in PSR are the SPX domain proteins, named for Suppressor of yeast gpa1 (ScSyg1), yeast Phosphatase 81 (ScPho81) and human Xenotropic and polytropic retrovirus receptor 1 (HsXPRI).

The first SPX containing proteins were discovered in budding yeast, the Pi exporter ScSyg1 and the nuclear cyclin dependent kinase inhibitor ScPho81 which regulates ScPho4 activation of PSR genes (Hamburger *et al.*, 2002; Wang *et al.*, 2004). SPX domains have been found N-terminally of various proteins in eukaryotes, typically involved with signal transduction (Barabote *et al.*, 2006). This hydrophilic domain is ~180 amino acids long and has had the speculated function of intracellular Pi sensor, which within the last decade has been more clearly defined as interaction with Pi, inositol polyphosphates (InsPs) and inositol pyrophosphates (PP-InsPs) (Wild *et al.*, 2016; Puga *et al.*, 2017).

4.1.1 SPX domains are Pi sensors

Through determination of SPX domain crystal structures from three proteins [ScVtc4, HsXPRI and *Chaetomium thermophilum* Gde1] (Wild *et al.*, 2016), a conserved SPX triple-helix bundle motif revealed two basic surface clusters suitable to interact with larger P-containing (phosphate-containing) ligands (Figure 4.1; Wild *et al.*, 2016). The Phosphate binding cluster (PBC) and Lysine surface cluster (LSC) contain highly conserved residues within eukaryotic SPX

domains, and together form an extended positively charged surface for ligand interaction (Wild *et al.*, 2016). Both CtGde1 and CtVtc4 structures were solved in complex with InsP₆ (PDB, 5IJJ and 5IJP), showing co-ordination of InsP₆ where the axial P2 is anchored within the PBC pocket (Wild *et al.*, 2016). The authors describe the high structural plasticity of the N-terminal binding site which may allow binding of different InsP/PP-InsP ligands and note that InsP binding stabilizes the N-terminal tail potentially conferring conformational changes.

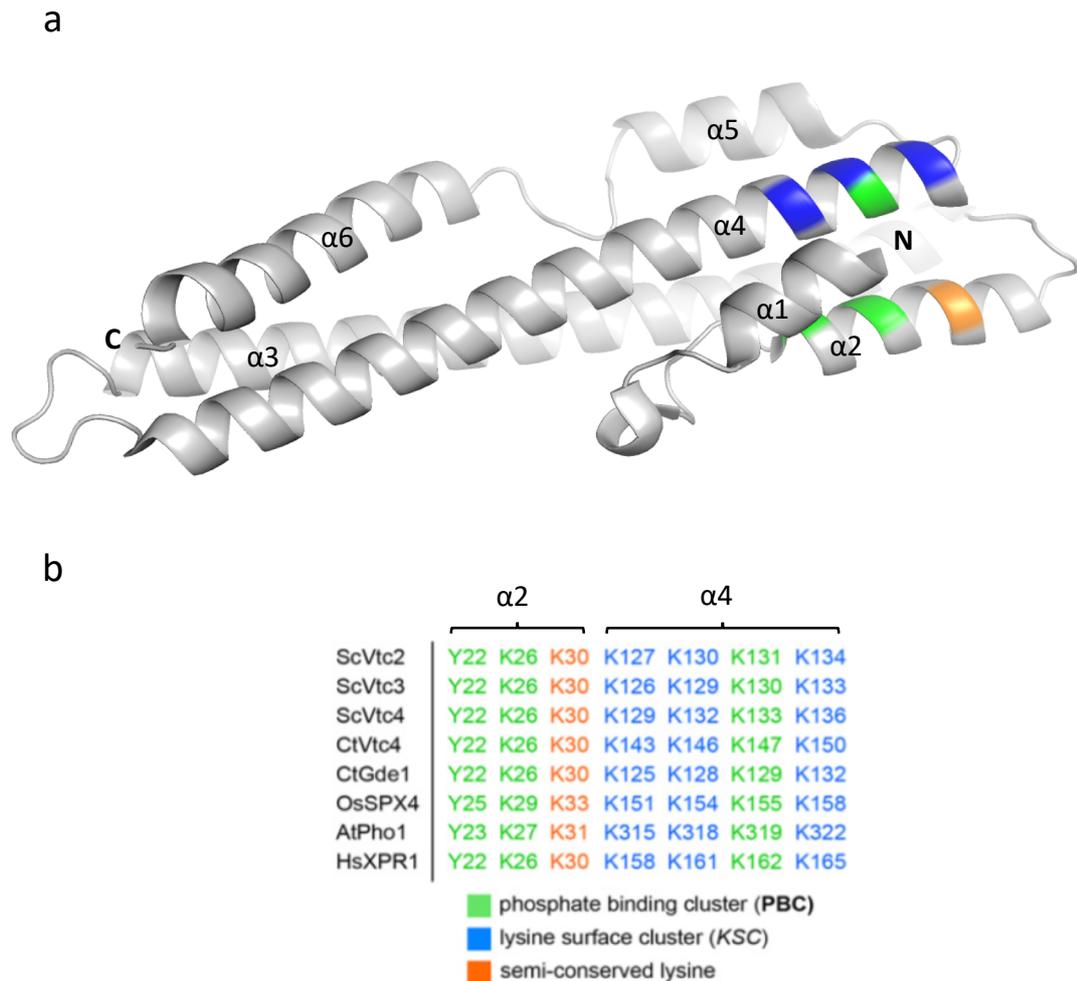


Figure 4.1 The SPX domain structural motif with conserved SPX fingerprint residues. (a) The N-terminal SPX domain (residues 1-179) from ScVtc4 (PDB, 5IIG) presented as cartoon (inspired by Jung *et al.*, 2019) with conserved residues from PBC and LSC highlighted according to table (b) Summary of conserved surface cluster residues from different SPX domain proteins (adapted from Wild *et al.*, 2016).

The P ligand binding affinity of yeast, animal and plant SPX-domains was explored by Wild *et al.*, (2016) using *in vitro* assays including nuclear magnetic

resonance (NMR), isothermal titration calorimetry (ITC) and microscale thermophoresis (MT). SPX domains were obtained through truncation of several SPX domain proteins and ligand binding affinities were compared through determination of equilibrium dissociation constants, K_d (Wild *et al.*, 2016). Pi binding was assessed through NMR, where binding affinity of SPX^{ScVtc2} and SPX^{HsXPRI} was determined in the millimolar range [approx. 5 and 23 mM respectively] (Wild *et al.*, 2016). ITC-determined K_d revealed higher affinity binding (10^4 to 10^6 M⁻¹) for the InsP₆ ligand of SPX^{ScVtc2}, SPX^{CtVtc4} and SPX^{CtGdel} (Wild *et al.*, 2016). Whilst, the K_d of SPX^{ScVtc2} binding of InsP₆ and 5-InsP₇ assessed by MT showed similar high affinity binding between InsPs (40 or 50 nM), the rice OsSPX4/OsPHR2 complex showed greater ligand discrimination with highest affinity for 5-InsP₇ (~7 μM), approx. seven times that of InsP₆ binding and over ten times that of InsP₅ (Wild *et al.*, 2016). While the different methods used do not allow direct comparison between the different proteins, it is possible that SPX protein binding partners can influence discrimination between different InsP ligands.

4.1.2 Characterization of plant SPX domain containing proteins

Plant SPX domain proteins have been characterised in species including: rice [*Oryza sativa*] (Wu *et al.*, 2013), soybean [*Glycine max*] (Yao *et al.*, 2014), oilseed rape [*Brassica napus*] (Du *et al.*, 2017), wheat [*Triticum aestivum*] (Kumar *et al.*, 2019) and *Arabidopsis thaliana* (Duan *et al.*, 2008). In *Arabidopsis thaliana*, there are twenty SPX domain proteins, which are categorized into four subfamilies/classes dependent on extra C-terminal domains as follows: (1) SPX, (2) SPX-RING (Really interesting new gene), (3) SPX-MFS (Major facilitator superfamily), (4) SPX-EXS (named for ScErd1, HsXpr1 and ScSyg1). Within each class, SPX domain proteins perform roles in regulating Pi homeostasis [Figure 4.2] (Jung *et al.*, 2018).

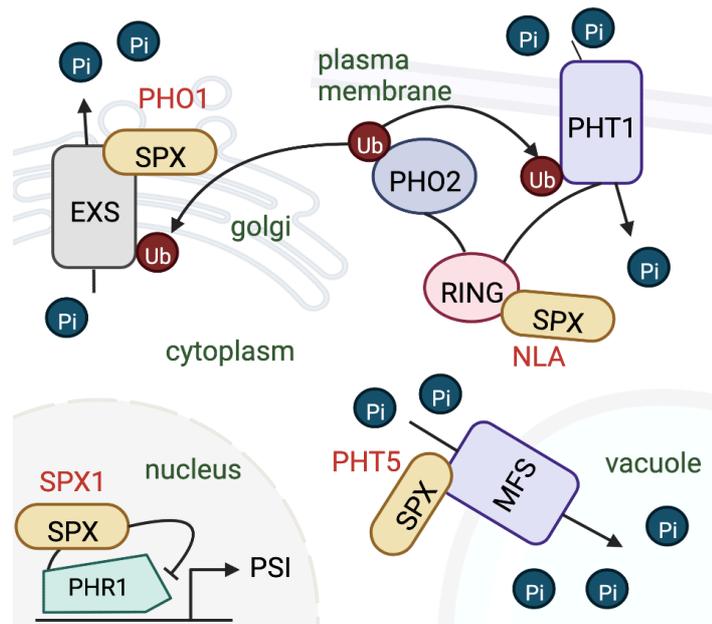


Figure 4.2 Schematic examples of plant SPX domain proteins with different C-terminal domains modulating Pi homeostasis; via PSI (Phosphate starvation inducible) gene regulation, Pi transport (SPX-MFS or PHT5, and SPX-EXS or PHO1) and mediation of degradation of Pi transporters (SPX-RING or NLA). Adapted from Wild *et al.* (2016), created with Biorender.com.

There are four Arabidopsis Class 1 SPX domain proteins, named AtSPX1-4, which exclusively contain the SPX domain and are commonly referred to as SPX proteins (Duan *et al.*, 2008). AtSPX1-3 genes are induced upon Pi starvation in roots and shoots and are under control of Phosphate starvation response (PHR) transcription factors which bind the PIBS *cis*-element [PHR1 binding site; GNATATNC] (Rubio *et al.*, 2001; Duan *et al.*, 2008). AtSPX1 and AtSPX2 localize to the nucleus and share functional redundancy in inhibition of PHR activity and PHR-independent transcription regulation (Duan *et al.*, 2008; Puga *et al.*, 2014). AtSPX3 localizes to the cytoplasm and is thought to sense intracellular Pi and negatively regulate AtSPX1, though this role has not been established (Duan *et al.*, 2008). AtSPX4 also functions as a regulator of PSR gene expression, however it localizes to the nucleus and cytoplasm and acts as a modulator of transcription factors in Pi-replete/low Pi conditions in shoots (Osorio *et al.*, 2019).

Class 2 SPX domain proteins contain the C-terminal RING domain, encoding ubiquitin E3 ligases involved in nitrate-dependent Pi homeostasis (Kant *et al.*, 2011). In Arabidopsis, the Nitrogen limitation adaptation 1 (NLAI) protein, also referred to as Benzoic acid hypersensitive 1 (BAH1), localizes in nuclear speckles and to the plasma membrane (Peng *et al.*, 2007; Lin *et al.*, 2013), in the latter case targeting Phosphate transporter 1 (PHT1) Pi transporters for degradation.

Class 3 SPX domain proteins contain the C-terminal MFS domain, in Arabidopsis referred to as the Phosphate transporter 5 (PHT5) subfamily or Vacuolar phosphate transporter (VPT). This subfamily consists of three tonoplast bound isoenzymes: AtPHT5;1, AtPHT5;2 and AtPHT5;3; whose members mediate Pi influx into the vacuole (Liu *et al.*, 2015). *PHT5* genes are expressed in distinct patterns with partial overlap, where *AtPHT5;1* is the most abundant transcript and is expressed in most tissues (Liu *et al.*, 2016). *AtPHT5;1* is upregulated in the roots under low Pi conditions (Liu *et al.*, 2016) and has been characterized as the predominant contributor to cytosol-to-vacuole Pi partitioning, important in maintenance of Pi homeostasis (Luan *et al.*, 2019).

Lastly, the Class 4 SPX domain proteins also referred to as the Phosphate 1 (PHO1) family proteins contain a C-terminal EXS domain. In Arabidopsis, this family consists of 11 proteins, which are Pi efflux carriers that load Pi into the xylem aiding root-to-shoot transport (Hamburger *et al.*, 2002; Wang *et al.*, 2004). *PHO1* is mainly expressed in the pericycle tissues of the root vascular cylinder and localizes to the trans-Golgi network, a network of membranes and associated vesicles (Arpat *et al.*, 2012).

4.1.3 AtSPX1 is an InsP_8 receptor

AtSPX1 is a Pi-sensitive accessory protein, which negatively regulates AtPHR1, master switch of PSR (Puga *et al.*, 2014). The AtSPX1-PHR1 interaction was first identified using yeast two-hybrid screening with truncated AtPHR1 (Δ PHR1, aa 208–362), and was shown to be Pi dependent *in vivo* through co-immunoprecipitation assays where protein-protein interaction was only detected

in transgenic plants (co-expressing GFP-SPX1 and HA-PHR1) grown in Pi sufficient conditions (Puga *et al.*, 2014).

It was first proposed that Pi was the signal molecule that would directly interact with the SPX domain of AtSPX1 and facilitate AtPHR1 binding (Puga *et al.*, 2014). Inhibition of AtPHR1 binding to the PIBS promoter was explored using Electrophoretic Mobility Shift Assays (EMSAs) with increasing additions of AtSPX1, in the presence and absence of Pi (Puga *et al.*, 2014). It was shown that AtSPX1 inhibits AtPHR1-DNA binding with dosage dependence of Pi concentration in the binding buffer. An optimum inhibition of AtPHR1-DNA binding in the presence of AtSPX1 was observed with the addition of 15 mM Pi and the half maximal inhibition concentration (IC₅₀) value of 0.3 mM Pi (Puga *et al.*, 2014).

In higher plants, the vacuole acts as a Pi storage such that in Pi sufficient conditions, they contain 85-95% of total Pi, whilst in Pi deficient condition, vacuolar Pi is mobilized to the cytosol and chloroplasts where Pi may be utilized (Bieleski, 1973). In Arabidopsis, cytosolic Pi has been quantified via ³¹P-NMR at 55-75 μM and shown to decrease rapidly upon external Pi deficiency (Pratt *et al.*, 2009). Pi accumulates at higher concentrations within plant organelles, such as plastids and mitochondria [approx. 7 mM Pi in Arabidopsis] (Pratt *et al.*, 2009), however Pi levels in nuclei have not been reported in Arabidopsis.

Following the Wild *et al.* (2016) study; a comparison of InsP₆ and Pi effect on AtSPX1 inhibition on AtPHR1-DNA binding was published. In this study, GST-tagged AtSPX1 was generated for surface plasmon resonance (SPR) studies. It was shown that the AtSPX1-AtPHR1 interaction occurs before AtPHR1-DNA binding, and cannot displace AtPHR1 from DNA (Qi *et al.*, 2017). Both the AtSPX1 and DNA binding site of AtPHR1 were therefore thought to be shared or in close proximity, which was later supported by a study identifying the AtPHR1 binding site within the coiled-coil domain (Qi *et al.*, 2017; Ried *et al.*, 2021). Upon addition of 5 mM Pi, a lower equilibrium of the AtPHR1-DNA binding signal was

observed, whereas a greater signal reduction was observed with the addition of 500 μM InsP_6 (Qi *et al.*, 2017), corroborating the finding of Wild *et al.* (2016) that SPX domains preferentially bind InsP_6 over Pi.

Most recently, Dong *et al.* (2019) have provided genetic and biochemical analyses proposing 1,5- InsP_8 as the true AtSPX1 ligand, acting as a proxy for intracellular Pi and co-ordinating the AtSPX1-AtPHR1 interaction (Figure 4.3). Levels of 1,5- InsP_8 were established to positively correlate with intracellular Pi concentration through polyacrylamide gel electrophoresis (PAGE) analysis (Dong *et al.*, 2019). Given that AtVIH1 and AtVIH2 produce 1,5- InsP_8 in Arabidopsis (Desai *et al.*, 2014), by investigating PSR in the *vih1 vih2* double mutant, which exhibits constitutive activation of PSI (phosphate starvation inducible) genes and Pi overaccumulation, the authors conclude that AtSPX1 inhibition of AtPHR1 is 1,5- InsP_8 dependent (Dong *et al.*, 2019). Co-immunoprecipitation assays with AtSPX1 and AtPHR1 show that under Pi deplete conditions there is no interaction, yet addition of 1 μM 1,5- InsP_8 and not 10 μM 5- InsP_7 can restore the AtSPX1-AtPHR1 interaction (Dong *et al.*, 2019).

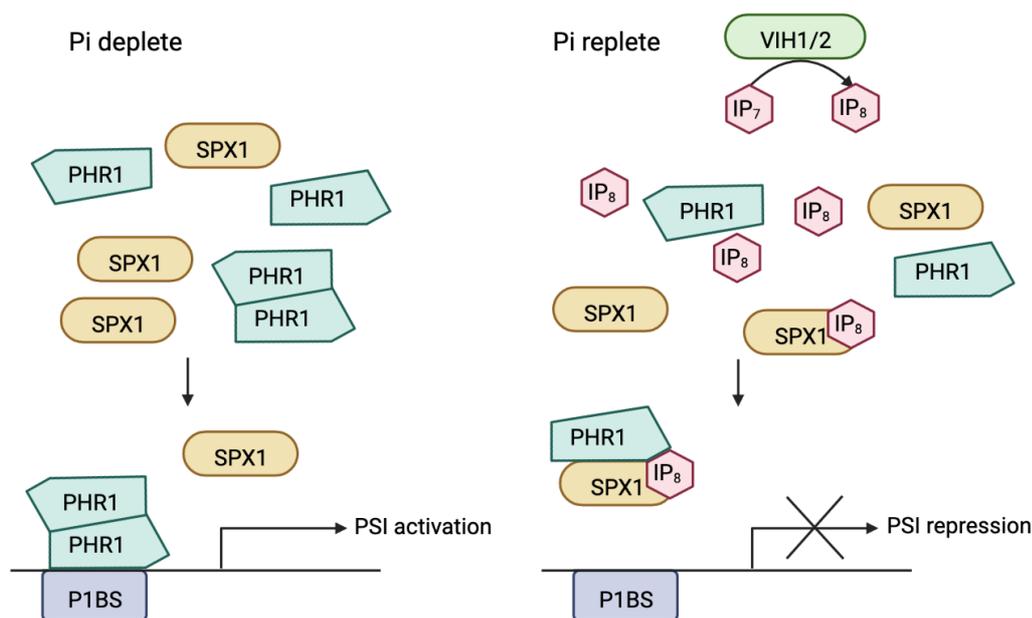


Figure 4.3 Model for AtSPX1 regulation of AtPHR1 PSR promoted by 1,5-InsP₈. Under Pi deplete conditions and in the absence of 1,5-InsP₈, constitutively expressed AtPHR1 is free to bind to the P1BS motif in promoter regions of PSI targets, including AtSPX1, and induce expression. Under Pi-replete condition in which Pi positively correlates to 1,5-InsP₈ availability, 1,5-InsP₈ promotes AtSPX1 sequestering AtPHR1, and inhibiting AtPHR1 dimerization and activation of PSI genes, whilst also providing a negative feedback loop of AtSPX1 expression. Adapted from Dong *et al.* (2019), created with Biorender.com.

Although 1,5-InsP₈ is the proposed ligand for Pi sensing in Arabidopsis, negatively regulating PSR, the roles of InsP₇ ligands or signalling molecules are not well understood. In determination of roles for InsPs and PP-InsPs, the abundance and compartmentalization of different molecules, as well as the ligand binding affinities to SPX domains (or other InsP binding proteins) should be considered. The reversibility of the enzymes involved and their nucleotide dependence are key parameters that have largely been ignored thus far (Whitfield *et al.*, 2020; Wang *et al.*, 2021).

4.1.4 Experimental aims

SPX domains can bind Pi, InsPs and PP-InsPs, which may facilitate protein-protein interactions or SPX protein extra-domain activity. Discrimination between different InsP species by Arabidopsis SPX proteins has not previously been described. In this study, we attempt to explore differences between P ligand binding of three Arabidopsis SPX domain proteins: AtSPX1, AtSPX3 and AtPHT5;1. Although AtSPX1 interaction with different P species has been demonstrated (Puga *et al.*, 2014; Dong *et al.*, 2019), quantitative binding data between several InsP/PP-InsP ligands has not been shown. Currently, there are no reports exploring P ligand binding of AtSPX3 or AtPHT5;1.

Whilst AtSPX1 and AtSPX3 are class 1 SPX proteins, which are strongly induced upon phosphate starvation (Duan *et al.*, 2008), they differentially localize to nucleus and cytoplasm, and exhibit low amino acid sequence identity of 46.5% [pBLAST (Altschul *et al.*, 1997)]. We therefore aimed to investigate whether SPX protein ligand specificity is determined by amino acid sequence. In contrast, AtPHT5;1 is tonoplast bound with a large C-terminal MFS domain and contributes to Pi allocation to the vacuole (Liu *et al.*, 2015). Vacuolar vs cytosolic Pi ratio is increased in *Atipkl* mutants and in the presence of IPK1 inhibitors (Liu *et al.*, 2016) we therefore aimed to investigate whether AtPHT5;1 activity is regulated by InsPs.

This investigation was approached by attempted cloning, expression and purification of the SPX domain proteins for *in vitro* studies. Fluorescence polarization (FP) displacement assays with the 2-FAM-InsP₅ probe were utilized to screen AtSPX1 against a set of InsPs and PP-InsPs, to provide a quantitative analysis of IC₅₀ values. Structural studies were sought through computational approaches from SPX protein sequences and crystallization trials were attempted, as an aid to comparisons between SPX domain proteins and ligand binding preferences.

During purification attempts of AtSPX1, a high A260/A280 ratio (measured on a NanoDrop™, Thermo Scientific) was observed. This prompted investigation into possible nucleotide-binding activity of AtSPX1, which has not been previously reported. All previous work has taken the view that the interaction of AtSPX1 with DNA involves only PHR and PIBS. Since 2-FAM-InsP₅ has been shown to be a powerful FP tool for study of InsP binding to diverse proteins and enzymes of InsP signalling (Watson *et al.*, 2016; Whitfield *et al.*, 2018, 2020), Fluorescein-tagged nucleotide, FAM-DNA, was used as a ligand for *in vitro* binding assays. As both DNA and InsPs/PP-InsPs are polyanionic with high charge density, the interplay of DNA and P ligands was similarly explored within FP displacement assays.

4.2 Materials and methods

4.2.1 Cloning and expression of SPX genes in *E. coli*

4.2.1.1 In-fusion cloning SPX genes into pOPINF

Clones of full length *AtSPX1*, *AtSPX3* and *AtPHT5;l* genes were kindly provided by Hui-Fen Kuo from Academia Sinica (Taipei, Taiwan), and were used as template for this work. Primers were designed to clone SPX genes into pOPINF (Table 4.1) via the In-fusion system (Takara Bio) according to manufacturer's guidelines.

Table 4-1 Primers designed for In-fusion cloning full length SPX domains; with regions for In-fusion cloning highlighted in purple italics.

Primer name	Primer sequence (5' to 3')	GC %	Tm °C
SPX1 FI	<i>AAGTTCTGTTTCAGGGCCCC</i> ATGAAGTTTGGTAAGAGTCTCA	36	57.6
SPX1 RI	ATGGTCTAGAAAGCTTTATTTGGCTTCTTGCTCAA	44	62.6
SPX3 FI	<i>AAGTTCTGTTTCAGGGCCCC</i> ATGAAGTTTGGAAAGAGGA	37	56.8
SPX3 RI	ATGGTCTAGAAAGCTTTATGGAATAGGAATCGGAGA	44	58.6
PHT5.1 FI	<i>AAGTTCTGTTTCAGGGCCCC</i> ATGGTGGCTTTTGGGAAATAC	43	55
PHT5.1 RI	ATGGTCTAGAAAGCTTTAATAGAGTGAGTTATAAGTACAACAAG	31	52

The In-fusion reactions were performed, and products were transformed into Stellar™ competent cells (Takara) with blue-white screening for positive selection, as detailed in section 3.2.2.2. The new constructs were propagated and isolated with the Wizard® Plus SV MiniPreps DNA Purification System (Promega), sequenced via the Mix2Seq kit (Eurofins) and finally transformed into competent expression *E. coli* strains including SHuffle® T7 Express (NEB), ArcticExpress™ RIL or RP (Agilent technologies) and BL21[DE3] (NEB). Glycerol stocks of each successful transformant cell line were made as described in section 3.2.2.3, and stored at -80 °C.

4.2.1.2 Expression trials and solubility testing of SPX gene products

Small scale expression trials for SPX protein production were conducted similarly as described in section 3.2.3.1. In addition to testing the expression of *AtPht5;l*

from pelleted cultures, solubility tests were performed as described in section 3.2.3.2 to isolate the soluble and insoluble fractions. These were analysed on an SDS-PAGE gel, described in section 2.2.1.8.

4.2.1.3 Large scale production of full length AtSPX1 and AtSPX3 in Arctic RIL

Glycerol stocks of ArcticExpress™ RIL cells containing *pOPINF:AtSPX1* and *pOPINF:AtSPX3* were used to prepare starter cultures of 300 mL LB with appropriate antibiotics (100 µg/mL ampicillin and gentamicin 50 µg/mL), growing overnight at 30 °C with shaking at 200 rpm. The cell cultures were then used to inoculate 20 mL into fresh 1 L LB media with antibiotics, which were grown for approximately 4 h at 30 °C, 200 rpm. The cultures were down tempered to 13 °C in pre-cooled incubators for 30-60 min before induction with 0.3 mM IPTG. Cells were left to express at 13 °C, 200 rpm for 24 h. Samples of expression cultures were taken and processed for SDS-PAGE analysis as described in section 2.2.1.8.

Cells were harvested as described in section 2.1.1.5, using resuspension buffer of 50 mM NaH₂PO₄ pH 6.5, 300 mM NaCl, 20 mM imidazole, 0.5% triton. Cell lysis was achieved using the French pressure cell as described in previous section 2.2.1.6, and SDS-PAGE analysis was performed to verify production of soluble protein.

4.2.1.4 Purification of SPX proteins

Purification proceeded as described in 2.2.1.7 methods section, excluding size exclusion chromatography and with the following differences. SDS-PAGE analysis was used to monitor protein yield and purity throughout the process. Final yield and purity were assessed using the NanoDrop™, where absorbance at 280 nm (A₂₈₀) was used to determine protein concentration (c) calculated using the Beer-Lambert equation: $A_{280} = c \times \epsilon \times b$ (ϵ is the wavelength-dependent protein extinction coefficient, b is the pathlength), and A₂₆₀/A₂₈₀ ratio to assess sample purity.

All binding and elution buffers for AtSPX1 and AtSPX3 purification were made with a pH of 6.5. During AtSPX1 purification, an addition of a high salt pre-wash with 50 mM NaH₂PO₄ pH 6.5, 1 M NaCl on the NiNTA HiLoad column was necessary to reduce nucleotide contamination. To reduce the high imidazole content in the eluate buffer from the NiNTA HiLoad column, the pooled samples were diluted with 50 mM NaH₂PO₄ pH 6.5 300 mM NaCl solution to approximately 20 mM imidazole, and then concentrated down to a small volume (approx. 5 mL) for cleavage. The 6xHis tag was cleaved overnight with HRC 3C protease (ThermoFisher) before using heparin chromatography. The binding and elution buffers for heparin chromatography had 10 mM sodium phosphate rather than 20 mM Tris and had additional 2 mM DTT and 5 % glycerol, to aid with correct protein folding and stability.

As a final step in the purification process of AtSPX1, buffer exchange was performed using PES membranes into a suitable buffer for protein stability instead of using size exclusion chromatography to maximise protein yield. Two buffer conditions were tested using the NanoDrop™ One (Thermo Scientific) to investigate whether AtSPX1 could tolerate low salt condition, using 10 mM sodium phosphate, pH 6.5, to dilute concentrated SPX1 from heparin chromatography in approximately 900 mM NaCl to 300 or 600 mM NaCl final concentration. The final buffer, which was subsequently used for in all AtSPX1 preparations was 10 mM sodium phosphate, pH 6.5, 600 mM NaCl, 5 mM DTT, 10% glycerol.

4.2.2 Cloning SPX domains into *Pichia pastoris*

4.2.2.1 Cloning SPX domains into pGAPZαA

To identify SPX domains for cloning from *AtSPX1*, *AtSPX3* and *AtPHT5;1*, prediction servers were used to determine the boundaries of the SPX domains. Firstly, nucleotide sequences were translated into amino acid sequences using ExpASY Translate tool. Protein sequences were used as query searches in InterProScan (Blum *et al.*, 2021) for domain annotation, Delta-BLAST (Boratyn *et al.*, 2012) for protein homology detection, PSIPRED (Buchan and Jones, 2019)

for secondary structure prediction and TMHMM (Krogh *et al.*, 2001) for predicting regions of disorder. Results from these searches helped define the SPX domains of interest as residues: 1-160 for AtSPX1, 1-147 for AtSPX3 and 1-156 for AtPHT5; hereafter designated as SPX^{AtSPX1}, SPX^{AtSPX3} and SPX^{PHT5.1}. These were then targeted for cloning into pGAPZ α A (Invitrogen) for expression trials in *Pichia pastoris*.

The pGAPZ α A vector for constitutive, secretory expression, alongside the *Pichia pastoris* strains of X-33 (wildtype) and KM71H (*aox1::ARG4, arg4*) were kindly provided by Raquel Rodriguez (UEA). These were used to make glycerol stocks (as described in 3.2.2.3.) for future work.

To aid two methods of cloning SPX domains into pGAPZ α A, In-fusion cloning sites and restriction enzyme sites were incorporated into primer design, such that each PCR product would include: Infusion site – *EcoRI* site – SPX domain – *XbaI* site – Infusion site. Primers pairs (Table 4.2) were synthesized by Eurofins. To amplify the insert for cloning, Phusion polymerase was used per manufacturer's guideline in 20 μ L reaction volume (4 μ L HF buffer, 0.4 μ L dNTPs (10 mM), 1 μ L 10 μ M F&R primers, 1 μ L 10 ng DNA, 0.2 μ L Phusion, 13.4 μ L water). Gradient PCR was initially used to find the optimum annealing temperature for each product, with following protocol: initial incubation of 98 °C for 30 s, then 30 cycles of denaturation at 98 °C for 10 s, annealing at 58-70 °C for 20 s and extension at 72 °C for 20 s, with final extension at 72 °C for 5 min. All PCR products were confirmed with gel electrophoresis on 1% agarose gels and purified with the Wizard[®] SV Gel and PCR Clean-Up System (Promega) kit according to manufacturer guidelines. Optimum annealing temperatures for SPX^{AtSPX3} and SPX^{AtPHT5.1} constructs (67 and 62 °C respectively) were subsequently used to provide materials for downstream experiments. PCR with the SPX^{AtSPX1} construct was unsuccessful under the conditions tested.

Table 4-2 Primers designed for two methods of cloning SPX domains into pGAPZ α A; where 5' infusion sites are designated purple, *EcoRI* sites are green, protein specific are red and *XbaI* sites are blue.

Primer name	Primer sequence (5' to 3')
SPX ^{SPXI} F	AGAGAGGCTGAAGCTGAATTCAGTTTGGTAAGAGTCTCAGCA
SPX ^{SPXI} R	GAGATGAGTTTTTGTCTAGAAAGTTGCTGAAGAAGCTTTCTGGATG
SPX ^{SPX3} F	AGAGAGGCTGAAGCTGAATTCAGTTTGGAAAGAGGATTAAGAAGACAGATAC
SPX ^{SPX3} R	GAGATGAGTTTTTGTCTAGAAAGTTGATGAAGAAGCTTTTGAATAAATGGTGATC
SPX ^{PHT5.1} F	AGAGAGGCTGAAGCTGAATTCGTGGCTTTTGGGAAATACTTG
SPX ^{PHT5.1} R	GAGATGAGTTTTTGTCTAGAGTTGAAGCTGAGAGTAAGGGTGATTAG

For traditional cloning, double restriction digests of each purified PCR product and pGAPZ α A vector were performed with ThermoFisher Fast Digest enzymes, *EcoRI* and *XbaI*, in 50 μ L reaction volume: 10 μ L (5 μ g) DNA template, 2.5 μ L each enzyme, 5 μ L 10x Fast digest buffer. Reactions were incubated at 37 °C for 18 min, then enzymes deactivated at 80 °C for 5 min. Restriction digest products were verified with DNA gel electrophoresis on 1 % agarose gels, gel extracted and purified with Wizard[®] SV Gel and PCR Clean-Up System (Promega) kit.

Ligation reactions with T4 DNA ligase (Promega) were set-up according to manufacturer guidelines on ice overnight, with multiple ratios of vector:insert (3:1, 1:1, 1:3, 1:5) in 15 μ L reaction volumes (containing 1.5 μ L 10x ligation buffer and 0.5 μ L ligase). Gel electrophoresis was used to assess successful ligation, and successful products were gel extracted and purified (as described above). Constructs were then heat-shock transformed into the *E. coli* plasmid propagation strain (NEB stable competent cells) as described in section 2.2.1.3, using low salt LB containing 25 μ g/mL zeocin for positive selection.

The alternative In-fusion method with purified PCR products of SPX domains and pGAPZ α A were conducted as described in section 3.2.2.2. Reaction products were purified using a PCR clean up kit and transformed into NEB stable competent cells using the same method for positive selection described above.

Successful cloning of SPX domains was verified with DNA sequencing. Mix2Seq (Eurofins) kits were used according to manufacturer's instruction, with sequencing primers shown in (Table 4.3). The sequence data was analysed using 4Peaks v1.8 (Nucleobytes) for validation, and Clustal Omega (Madeira *et al.*, 2019) to compare with expected sequence of gene.

Table 4-3 Sequencing primers for validation of SPX domain cloning.

Primer name	Primer sequence (5' to 3')
pGAP F	GTCCCTATTTCAATCAATTGAA
3'AOXI	GCAAATGGCATTCTGACATCC

In preparation for *Pichia* transformations, MIDI preps of successful cloned strains were made to generate 5-10 µg of each pGAPZαA construct. These were linearized with *AvrII* (NEB) overnight at 4 °C ready for *Pichia* transformation reactions.

4.2.2.2 Preparing competent *Pichia pastoris*

Pichia competent cells were prepared by first growing 50 mL cultures of X-33 and KM71H overnight, at 30 °C and 200 rpm, in YPD media (1% yeast, 2% peptone, 2% dextrose) with 100 µg/mL kanamycin.

Overnight cultures were pelleted in 50 mL Corning tubes at 4000 rpm for 5 min, media was discarded. From this point, all centrifugation steps were performed at 4000 rpm, 5 min and 4 °C. Pellets were suspended in 20 mL ice-cold sterile water, then topped up to 50 mL and centrifuged. Water was discarded and the cells were rested on ice, whilst fresh SED solution (1 M sorbitol, 50 mM TrisHCl pH 7.5, 20 mM DTT, 25 mM EDTA pH 8.0) was made. Pellets were suspended in 25 mL SED solution and incubated at rt for 10-15 min before centrifugation. Excess solution was discarded, and cells then resuspended in 20 mL ice cold 1 M sorbitol, pelleted and solution discarded again. Finally, cells were suspended in 3 mL of 1 M sorbitol and kept on ice to be used immediately for transformations.

4.2.2.3 Transformation of SPX constructs into *Pichia pastoris*

For each transformation reaction, 10 μL linearized recombinant plasmid was added to 390 μL competent *Pichia* cells into pre-chilled 0.2 cm electroporation cuvettes (Bio-Rad), and kept on ice for 10 min.

The cell suspensions were electroporated on the fungi/Pic setting of the micropulser (Bio-Rad), and then 1 mL of 1 M sorbitol was added into the cuvette. The transformed cells were transferred to a sterile 15 mL tube and left to recover at 30 °C for 24 h. Different volumes (10, 25, 50, 100 and 200 μL) were spread onto YPDS media plates (1% yeast, 2% peptone, 2% dextrose, 1 M sorbitol, 2% agarose) containing 100 $\mu\text{g}/\text{ml}$ zeocin™ for selection, and left to grow for 2-4 days at 30 °C. For each construct, six colonies were re-streaked onto YPD agar plates, containing 100 $\mu\text{g}/\text{ml}$ zeocin™, and left to grow at 30 °C for a further 2-4 days. These cell lines were used to conduct expression trials.

4.2.2.4 Small scale expression trials in *Pichia*

For each construct in X-33 and KM71H, six transformants from re-streaked plates and negative controls of untransformed X-33 and KM71H, were inoculated with a sterile loop into 2 mL of YPD media (in 24 deep well plates with air-tight lid cover) with no antibiotics. Cultures were grown over four days at 30 °C with shaking at 200 rpm and samples of 20 μL were taken (and 20 μL of YPD media replaced) at time points: 0, 24, 48, 72 and 96 h and prepared for SDS-PAGE analysis (section 2.2.1.8) of total protein profiles. No overexpressed protein was detected on this scale.

4.2.3 Characterization of SPX proteins

4.2.3.1 FP with AtSPX1

FP assays with AtSPX1 and analyses were performed as described in 2.2.2.2 with the following differences. The binding buffer used was 20 mM HEPES pH 6.5, 1 mM MgCl_2 with variant NaCl (0, 50 and 100 mM NaCl) as stated in the text. The ligands used to generate standard curves were ds FAM-oligomers (FAM-AT 40mer, FAM-GC 40mer, FAM-12mer, PIBS, PIBSr and 4x PIBS; Table 4.4)

annealed (as detailed in section 2.2.1) and 2-FAM-InsP₅ (Appendix 1). Binding assays were also performed with and without the addition of Pi (0, 5, or 10 mM).

Table 4-4 5'-FAM-labelled double stranded oligomers used in FP with AtSPX1.

Oligomer name	Nucleotide sequence (5' to 3')
PIBS	GAATTGAATATGCAATG
PIBSr	AATGACATGATGTAATG
4x PIBS	GAATTGAATATGCAATGGAATATGCTTAGGCATATTCATAGA ATATTCCTAGA
4x PIBSr	ATTAAGGAATCTATTAAGTATGAAGTGGAATGTACGCCCATTA ATTGAAACTTC
FAM-GC 40mer	CCCCGGGTACCGAGCTCGAATTCCTGGCCGTCGTTTTAC
FAM-AT 40mer	CCCCGGCGAATTAATTCGAATTCCTAAACGTAGTTTTAC
FAM-12mer	CGAATTAATTCG

Displacement assays (for more details see section 3.2.7.2) were performed with either 300 nM of AtSPX1, 2 nM of probe (as detailed above), 100 mM NaCl and the addition of 0 to 1 μ M of one of the following InsP ligands: InsP₆, Ins(1,4,5,6)P₄, Ins(3,4,5,6)P₄, Ins(1,3,4,5,6)P₅, and PP-InsP ligands shown in Table 4.5. This method was also used to displace 500 nM of AtSPX1 binding 2 nM FAM-InsP₅ probe with unlabelled oligomer (GC 40mer) titrations.

Table 4-5 Inositol pyrophosphate ligands used in FP

Abbreviation	Full description	
5-PP-InsP₄	5-diphospho- <i>myo</i> -inositol 1,3,4,6-tetrakisphosphate	^P
1-InsP₇	1D-1-diphospho- <i>myo</i> -inositol 2,3,4,5,6-pentakisphosphate	^P
3-InsP₇	1D-3-diphospho- <i>myo</i> -inositol 1,2,4,5,6-pentakisphosphate	^P
5-InsP₇	5-diphospho- <i>myo</i> -inositol 1,2,3,4,6-pentakisphosphate	^P
1,5-InsP₈	1D-1,5-bis-diphospho- <i>myo</i> -inositol 2,3,4,6-tetrakisphosphate	^P
3,5-InsP₈	1D-3,5-bis-diphospho- <i>myo</i> -inositol 1,2,4,6-tetrakisphosphate	^J

^P from Barry Potter; ^J from Henning Jessen.

4.2.3.2 Crystallography screens with AtSPX1

Freshly purified full length AtSPX1 (6 mg/mL) was centrifuged at 13,000 rpm for 10 min at 4 °C in preparation for crystallization trials. Crystallography screens were arranged with AtSPX1 using the sitting drop vapour diffusion method as

described in section 3.2.1.1 with all six commercial screens used before. Unfortunately, temperature control issues were experienced in the 4 and 16 °C cold rooms used to conduct screens due to a heatwave. No crystals grew, there was no time to repeat the experiment.

4.2.3.3 F-EMSA with AtSPX1

F-EMSA (fluorescence-based electrophoretic mobility shift assay) was attempted with AtSPX1 as described in section 2.2.2.3. However, the final binding buffer used was 20 mM HEPES pH 6.5, 1 mM MgCl₂, 0.05 mg/mL BSA, 50 mM NaCl and 1 mM EDTA.

4.2.3.4 Modelling SPX domain proteins

Structural theoretical models of AtSPX1, AtSPX3 and AtPHT5;l were made using the Phyre2 server (Kelley and Sternberg, 2009). Phyre2 uses the Hidden Markov Method to generate alignments of the query protein sequence against published protein structures. These alignments were used to produce homology-based models to predict tertiary structures.

4.3 Results

4.3.1 Cloning, expression and purification of SPX proteins

4.3.1.1 Cloning and expression of SPX domains in *E. coli*

To enable characterization of SPX domains in Arabidopsis, three proteins of interest were chosen (AtSPX1, AtSPX3 and AtPHT5.1) and cloned into the pOPINF vector for heterologous gene expression (for example see Appendix 12). Several *E. coli* expression strains (BL21, Rosetta2, ArcticExpress™, SHuffle®) were tested as hosts for protein production. Predicted protein sizes from ProtParam (ExPASy) were used to determine the success of protein production in each case, using SDS-PAGE analysis. Out of the three SPX domain proteins of interest, full length AtSPX1 (approx. 30 kDa) was produced for further investigation as discussed in more detail below.

Although apparent AtSPX3 (approx. 28 kDa) production was observed in both small and large-scale expression trials with ArcticExpress™ and Rosetta2 strains (Table 4.6), the protein of interest was largely lost in the first step of purification (Appendix 13). Extensive attempts were also made to produce the relatively larger protein AtPHT5.1 (~ 78 kDa) but these proved unsuccessful with the expression strains tested (Table 4.6; Appendix Figure 13.2), that is, despite successful cloning of the gene into pOPINF, as verified by DNA sequencing (Appendix 14).

Table 4-6 Summary of expression data of *pOPINF[T7lacO]-SPX* constructs in different *E. coli* strains

Construct	Strain	[IPTG] (mM)	Time	Expression	Solubility
AtSPX1	ArcticExpress™ RIL	0.1, 0.5	6h/ON*	++	+
	BL21 (DE3)	0.5	6h	-	N/A
	Rosetta 2 (DE3)	0.5	6h/ON	++	+
	SHuffle T7 Express	0.1, 0.5	6h/ON	+	+
	SHuffle T7 LysY	0.1, 0.5	6h/ON	+	+
AtSPX3	ArcticExpress™ RIL	0.5	6h/ON*	++	+
	BL21 (DE3)	0.5	6h	-	N/A
	Rosetta 2 (DE3)	0.5	6h/ON	++	+
	SHuffle T7 Express	0.5	6h	-	N/A
	SHuffle T7 LysY	0.5	6h	-	N/A
AtPHT5;1	ArcticExpress™ RP	1.0	ON	-	N/A
	Rosetta (DE3)	0.2	ON	-	N/A
	SHuffle T7 Express	0.2	ON	-	N/A
	SHuffle T7 LysY	0.2	ON	-	N/A

- is no expression/solubility

+ is expression/solubility detectable

++ is high expression/highly soluble protein.

Expression cultures were induced with IPTG at rt or 13 °C (*) and data collected from SDS-PAGE analysis.

4.3.1.2 Cloning and expression of SPX domains in *Pichia pastoris*

As an alternative to characterization of full-length SPX proteins, SPX domains from *AtSPX1*, *AtSPX3* and *AtPHT5;1* were identified and cloned into pGAPZαA for expression in *Pichia pastoris*. *P. pastoris* are methylotrophic yeast (in the absence of glucose they will use methanol as a carbon source), which are commonly used in the expression of recombinant proteins (Karbalaei *et al.*, 2020). In contrast to *E. coli* as an expression system, *P. pastoris* can assist correct protein folding (via the endoplasmic reticulum). The pGAPZαA vector (Invitrogen), which contains the GAP promoter for constitutive gene expression, targeted integration into the host genome, zeocin resistance and secreted expression, was chosen for this latter feature to allow for easier downstream purification processes. Successful cloning of SPX^{AtSPX3} and SPX^{AtPHT5;1} in pGAPZαA was confirmed in sequencing data (Appendix 15). However, this work was not completed due to time restraints, at the point of expression trials with SPX^{AtSPX3} and SPX^{AtPHT5.1}.

4.3.1.3 Large scale production of AtSPX1 in *E. coli*

AtSPX1 was successfully cloned into pOPINF and expressed in ArcticExpress™ RIL cells (Agilent). ArcticExpress™ cells allow low temperature cultivation to increase the recovery of soluble proteins, engineered with cold adapted chaperonins, Cpn10 and Cpn60 from *Oleispira antarctica*, to facilitate refolding activity at temperatures 4-13 °C. This expression system was used in attempting large scale culture (12 L) protein production.

Protein purification of AtSPX1 was performed using nickel affinity chromatography as a first step (Figure 4.4a; Appendix Figure A16.1), followed by His-tag removal, heparin chromatography (Figure 4.4b; Appendix Figure A16.2) and buffer exchange. Heparin chromatography was introduced during work up of purification, as it proved an essential step in the purification process to reduce the A260/280 ratio indicative of nucleotide contamination. Alone this step was insufficient, in itself, to remove nucleotide contamination (A260/280 of 1.24) and an additional high salt (1 M NaCl) pre-wash on the nickel affinity column was necessary to achieve a low A260/280 preparation (0.76). To investigate potential differences between high and low A260/280 preparations of protein, in each case the preparations were buffer exchanged into 20 mM TrisHCl pH 6.5, 200 mM NaCl, 2 mM DTT.

Even so, scaling up AtSPX1 production using the above as a final buffer proved challenging, due to protein precipitation at higher AtSPX1 concentrations. In an attempt to improve the protein stability, buffers containing higher NaCl levels were trialled in 10 mM sodium phosphate pH 6.5, 5 mM DTT and 10% glycerol. Whilst comparing precipitation using the NanoDrop™ between 300 and 600 mM in the final buffer, the latter (the approx. concentration at which AtSPX1 eluted from the heparin column) was sufficient to reduce protein losses. Using this final pipeline of purification, approx. 0.5 mL of 6 mg/mL of AtSPX1 was yielded from 12 L expression cultures, and SDS-PAGE analysis showed the final prep to have few protein contaminants (Figure 4.4c).

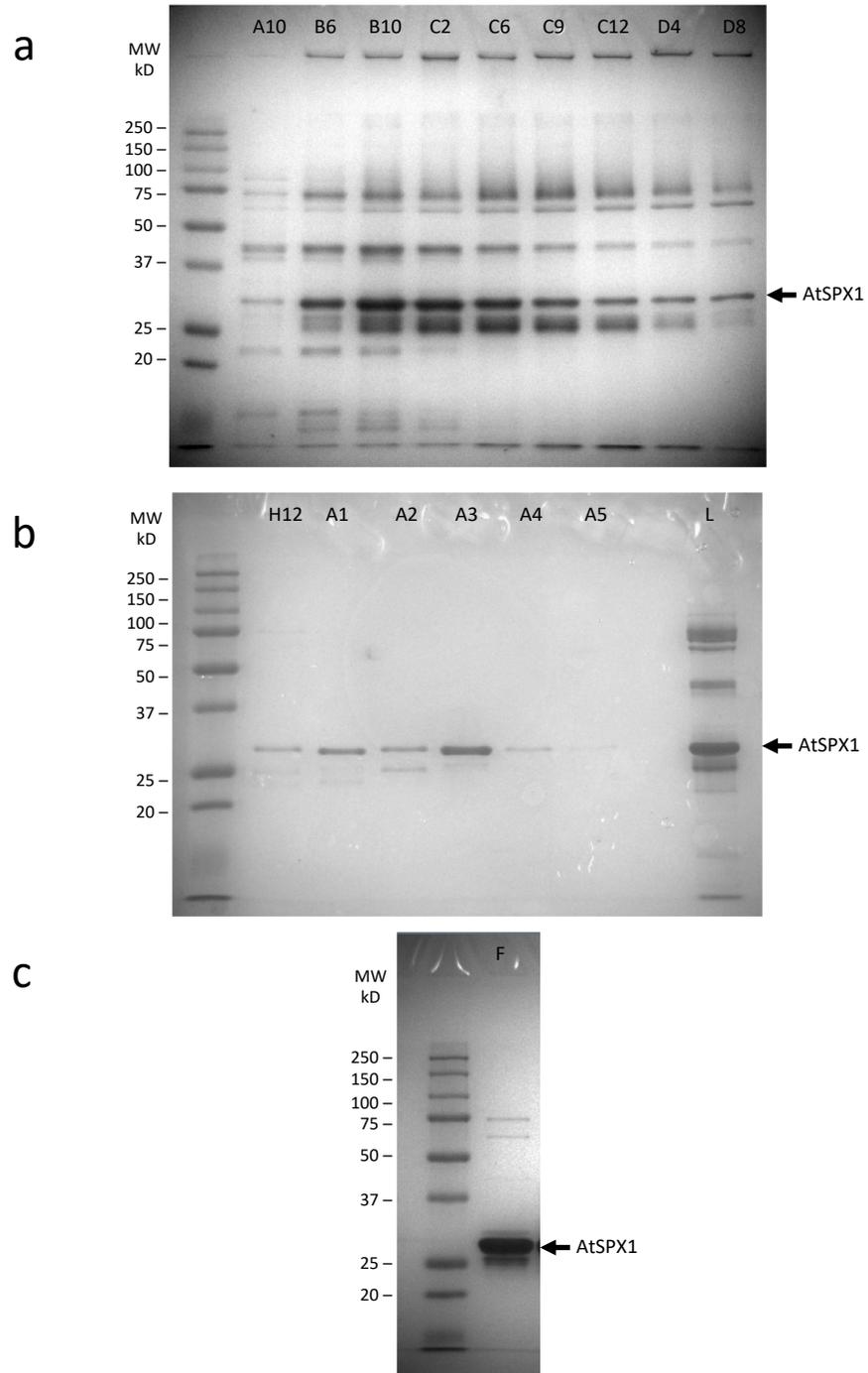


Figure 4.4 SDS-PAGE analysis of AtSPX1 (expected size of ~30 kDa) on 12% polyacrylamide gels, after different purification steps including: (a) fractions collected from nickel chromatography (A10-D8); (b) fractions collected from heparin chromatography (H12-A5) protein loaded onto heparin column (L); and finally, (c) purified AtSPX1 after buffer exchange (F), overloaded onto SDS gel.

4.3.2 Binding of InsPs and PP-InsPs by AtSPX1

4.3.2.1 Investigation of AtSPX1 ligand binding with FP

AtSPX1 has been reported to bind InsP and PP-InsP ligands (Wild *et al.*, 2016), with little understanding of binding preference. To further investigate AtSPX1 ligand binding, we employed the technique of fluorescence polarization (FP), first establishing 2-FAM-InsP₅ as a suitable fluorescent probe for AtSPX1. In both low and high A_{260/280} ratio preps of AtSPX1, the protein was shown to bind 2-FAM-InsP₅ despite neither binding curve reaching full saturation at the top protein concentration of 1 μ M AtSPX1 (Figure 4.5a). The effect of NaCl on the AtSPX1-2-FAM-InsP₅ interaction was explored with FP (Appendix 17), revealing that the greatest polarization dynamic ranges are observed at 0 mM NaCl and changes are only observed up to \sim 300 mM NaCl. To mitigate non-specific interactions whilst maintaining a large dynamic range of polarization values, 100 mM NaCl was chosen for use in further FP assays.

Due to the high A_{260/280} ratio observed during the purification of AtSPX1, possible nucleotide binding activity of AtSPX1 was initially explored by setting up FP binding assays with AtSPX1 and FAM-DNA. High and low A_{260/280} ratio preps of AtSPX1 were compared side by side, showing a stark difference where interaction with the DNA probe was only detected using the low A_{260/280} ratio AtSPX1 prep (Figure 4.5b). This exciting result prompted further exploration of AtSPX1 DNA binding activity, as detailed in section 4.3.3.

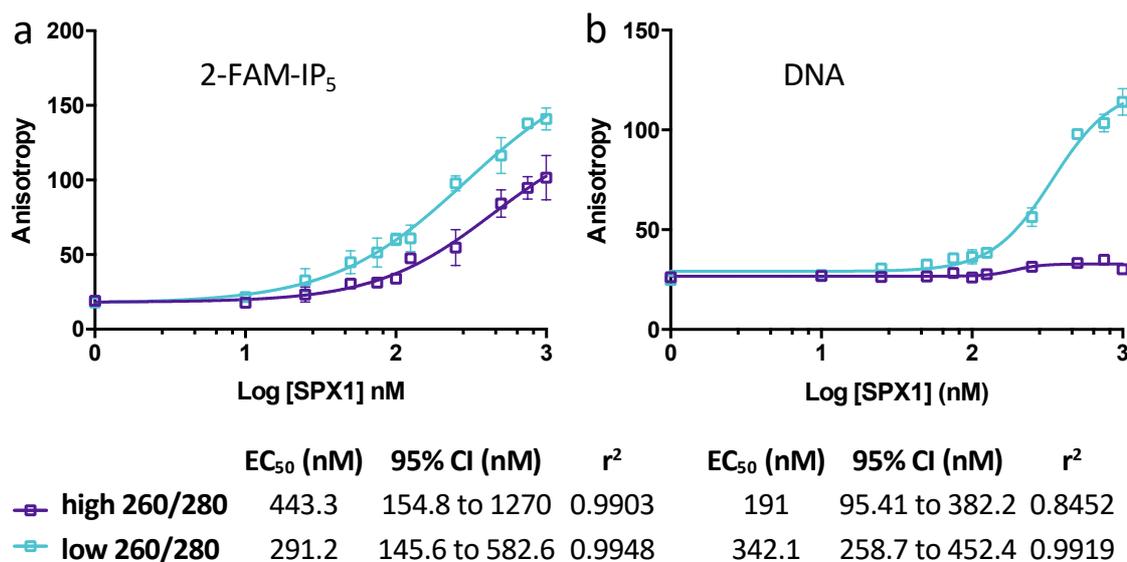


Figure 4.5 FP with high and low A260/280 ratio preps of AtSPX1 binding. Showing (a) 2-FAM-InsP₅ (2 nM) and (b) FAM-labelled DNA (2 nM) in 20 mM HEPES pH 6.5, 1 mM MgCl₂, 100 mM NaCl. Variable slope curves generated by Prism v6.0 (Graphpad) showing the mean \pm SD of fluorescence polarization data.

4.3.2.2 Investigation of AtSPX1 binding preference of InsPs and PP-InsPs

A suite of InsPs and PP-InsPs provided by Barry Potter, University of Oxford, and Henning Jessen, University of Freiburg, were assessed (Figure 4.6) for their ability to displace 2-FAM-InsP₅ in complex with AtSPX1 using FP and where possible obtain IC₅₀ values. Poor displacement of 2-FAM-InsP₅ from AtSPX1 was observed for the lower InsP species tested, including InsP_{4s} and InsP₅ (Figure 4.7a). Although InsP₅ appears to displace 2-FAM-InsP₅ at higher concentrations (above 100 nM), the inhibition curve did not saturate sufficiently to obtain a fit suitable to generate an IC₅₀ value. The 5PP-Ins(1,3,4,6)P₄ [5PP-InsP₄] was shown to displace 2-FAM-InsP₅ with a similar affinity to InsP₆ (IC₅₀s of 82 nM, CI respectively), where both ligands have 6 phosphates present. The more highly phosphorylated IPP-Ins(2,3,4,5,6)P₅ [1-InsP₇], 3PP-Ins(1,2,4,5,6)P₅ [3-InsP₇], 5PP-Ins(1,2,3,4,6)P₅ [5-InsP₇], 1,5PP-Ins(2,3,4,6)P₄ [1,5-InsP₈] and 3,5PP-Ins(1,2,4,6)P₄ [3,5-InsP₈] displayed stronger displacement of 2-FAM-InsP₅ than InsP₆ or 5PP-InsP₄ (Figure 4.7b), and similar IC₅₀ values (approx. 50-60 nM) were obtained for all of the InsP₇s and InsP₈s, making it difficult to distinguish specificity of AtSPX1 binding. It would therefore appear that AtSPX1 prefers interacting with

more phosphorylated InsP species but does not greatly discriminate between InsP₇ and InsP₈.

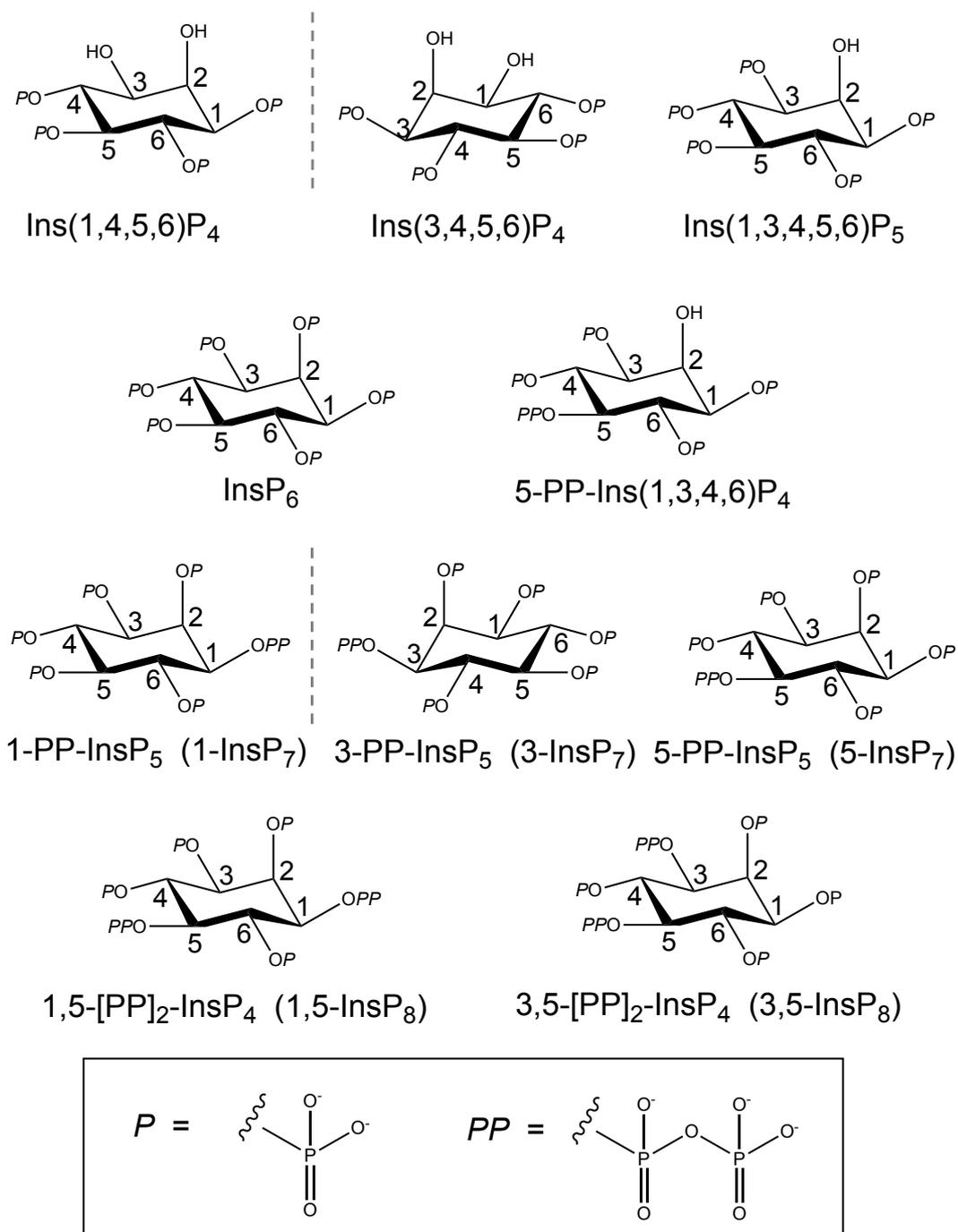


Figure 4.6 Inositol polyphosphates and inositol pyrophosphates used in this study. Structures rendered in ChemDraw.

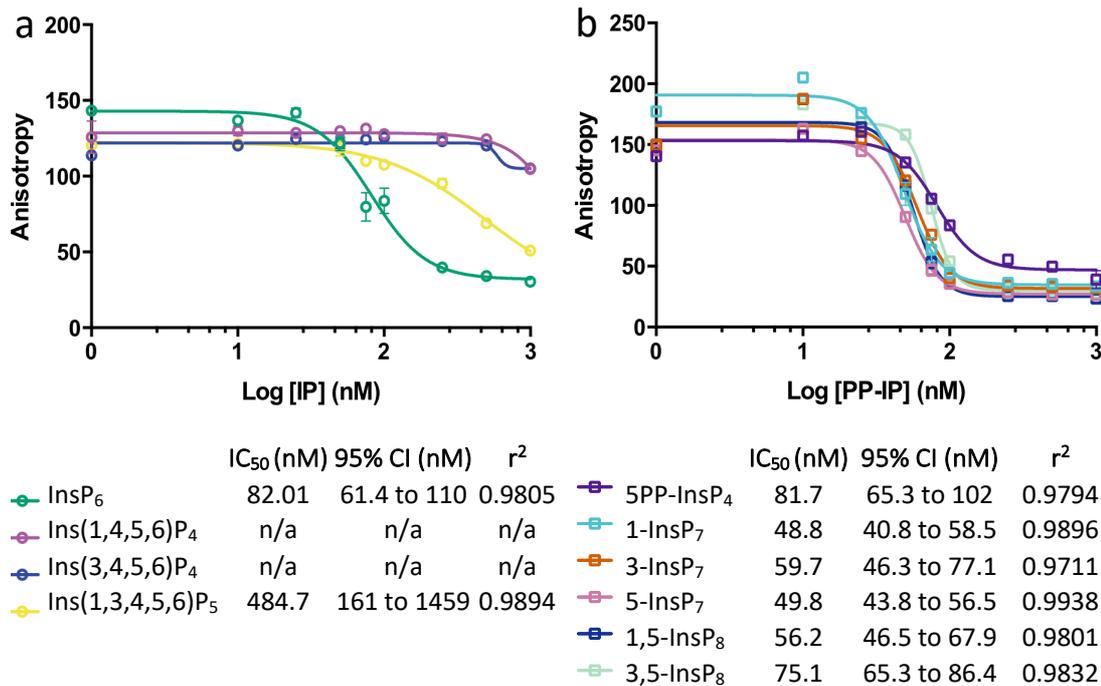


Figure 4.7 Displacement of 300 nM AtSPX1 binding of 2 nM 2-FAM-InsP₅ with (a) InsPs and (b) PP-InsPs. Variable slope curves generated by Prism v6.0 (Graphpad) showing the mean \pm SD of fluorescence polarization data. These binding assays were performed in 20 mM HEPES pH 6.5, 1 mM MgCl₂, 100 mM NaCl and incubated for 1 h before plate readings at 25 °C.

To test whether additions of DNA would displace AtSPX1-2-FAM-InsP₅ interaction, unlabelled DNA was titrated into the binding reaction (Appendix 18). Displacement of 2-FAM-InsP₅ was observed at relatively high DNA concentrations (above 100 nM) with IC₅₀ of approx. 3 μ M. This suggested that there may be a reciprocal interaction between InsP and DNA binding, which led to further investigation in section 4.3.3.2.

4.3.3 Interaction of AtSPX1 with DNA

4.3.3.1 AtSPX1 binds different FAM-labelled DNA probes

FP was used to monitor AtSPX1-DNA interactions *in vitro*, using different FAM-labelled oligomers (Figure 4.8). The binding site of PHR1, the PIBS promoter (17-nt), was chosen as a probe to monitor DNA binding, along with a randomised PIBS probe (PIBSr) and a 4x repeat of the PIBS probe (4x PIBS). The PIBS probe produced the tightest saturation curve, and relatively low dissociation constant

($K_d \sim 586$ nM). In comparison, the apparent tight binding observed from the FAM-12mer was subject to large error, the curve produced did not have sufficient anisotropy dynamic range to be confident of this fit.

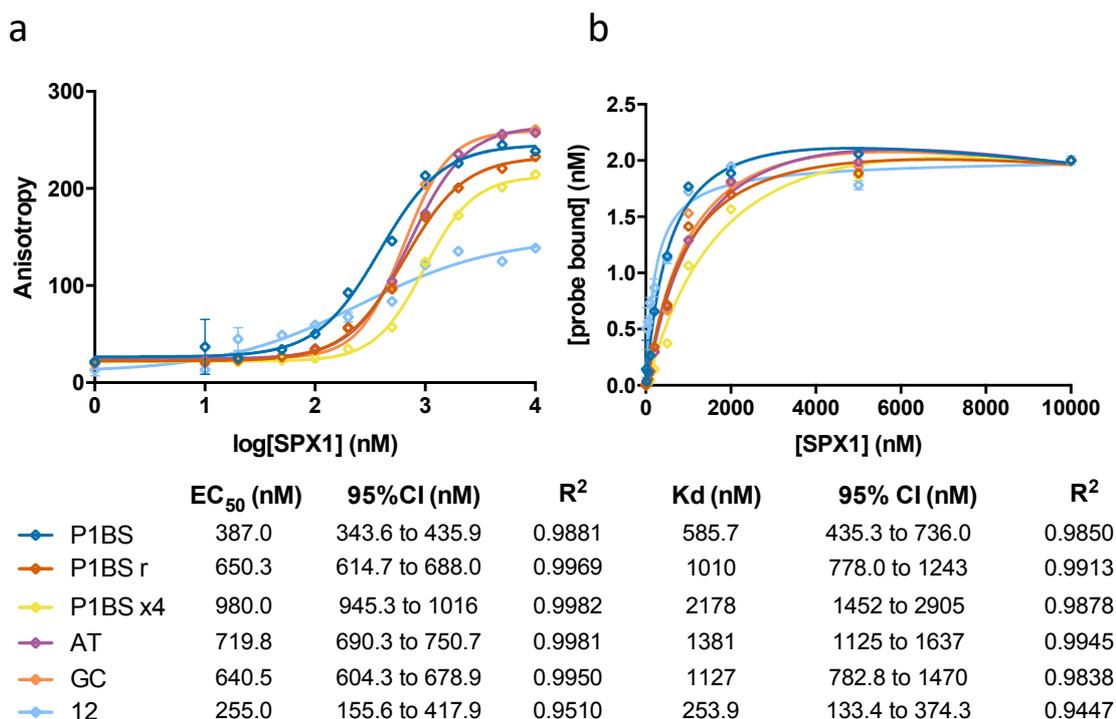


Figure 4.8 AtSPX1 (0-10 μ M) binding different 5' FAM-labelled oligomers (2 nM) in binding buffer (20 mM HEPES pH 6.5, 1 mM $MgCl_2$, 100 mM NaCl), monitored by FP. Variable slope curves generated by Prism v6.0 (Graphpad) showing the mean \pm SD of anisotropy data. These binding assays were incubated for 1 h before plate readings at 25 $^\circ$ C.

Direct comparison of the 17-nt P1BS binding element against P1BSr indicates an approximately 2-fold higher affinity to P1BS, which could suggest sequence specificity considering that the sequences contain the same AT/GC ratio and are the same length. The 4x P1BS repeat element with 54-nt has a much higher K_d (approx. 2 μ M) which may indicate that P1BS is not a consensus binding site of AtSPX1. When this assay was repeated, similar results were obtained for the P1BS and P1BSr probes (Appendix 19), and an additional randomised 4x P1BS (4x P1BSr) probe was included for analysis. The randomised 54-nt probe had a lower binding affinity compared to the 4x P1BS probe, further suggesting some sequence preference for P1BS given the presence of the same nucleotide content.

In contrast, AtSPX1 did not appear to have a preference between AT or GC 40mer sequences, binding both probes with a K_d of approximately 1 μ M.

Alternative monitoring of AtSPX1-DNA interaction was attempted using F-EMSA, as in previous chapters. However, amongst the conditions tested, the majority of AtSPX1 did not enter the NATIVE gel (Appendix 20). Despite this, a shift of the FAM-DNA can be observed with additions of AtSPX1.

4.3.3.2 AtSPX1-DNA-P ligand interactions

We may assume that AtSPX1, like other SPX proteins, encounters a spectrum of competing ligands in its intracellular context. No single InsP-ligand is likely to be present in the absence of others, and in the context of Pi homeostasis all will 'operate' in cellular environments with excess inorganic phosphate (estimated at 55 μ M to 7 mM, Pratt *et al*, 2009). The observed AtSPX1-DNA interaction with PIBS appeared to be strengthened upon addition of Pi (5 and 10 mM), demonstrated by the left shift of the saturation curves (Appendix 21) and ~1.5-fold decrease in K_d . This opposite effect was observed for AtSPX1-2-FAM-InsP₅ binding, in which additions of Pi appeared to lower polarization values (Appendix 22). Upon further investigation, it became clear that these observations were not due to the additional Pi, but the change in pH of the binding buffer as a result of the presence of higher levels of Pi (Appendix 22). Whilst adjusting the buffer over a range from pH 5.2-7.0 and not adding Pi, polarization values were higher at lower pH and the inverse was observed for AtSPX1-2-FAM-InsP₅ binding.

To examine the interplay of AtSPX1 interaction with DNA and InsP ligands, FP was used (Figure 4.9). The AtSPX1-DNA displacement by InsPs and PP-InsPs was notably different to that of 2-FAM-InsP₅, where phosphate number appeared to determine displacement strength. Firstly, InsP₆ demonstrated a several fold more effective inhibition of the AtSPX1-DNA interaction (IC_{50} ~ 138 nM) compared to the similarly phosphorylated PP-InsP counterpart (5PP-InsP₄, IC_{50} ~740 nM). Despite the fit of these curves not being optimal, in both cases the anisotropy

value drops to the baseline at the highest InsP concentration. Secondly, although generally the more highly phosphorylated InsPs and PP-InsPs demonstrate greater efficacy at DNA displacement, as was the case in the 2-FAM-InsP₅ displacement assays above, some distinctive patterns can be inferred. For example, the pyrophosphate bond in the C1 position appears to improve PP-InsP binding affinity in contrast to the pyrophosphate bond in the C3 position; this difference is best shown from the InsP₇ data where 1-InsP₇ was approx. 2x better at displacement than 3-InsP₇ (IC₅₀ 70 vs 140 nM). However, this difference is not statistically significant. This is further supported by the comparison between 1,5-InsP₈ and 3,5-InsP₈, where the 3,5-InsP₈ is less effective at displacing the DNA probe. Here, the compounds are identical other than the presence of the PP on the 1 or 3 position so are directly comparable. A higher resolution data set covering more points between 25 and 250 nM ligand would be helpful in distinguishing displacement curves and understanding InsP ligand discrimination. However, due to limited supplies of PP-InsPs in particular, these experiments could not be performed.

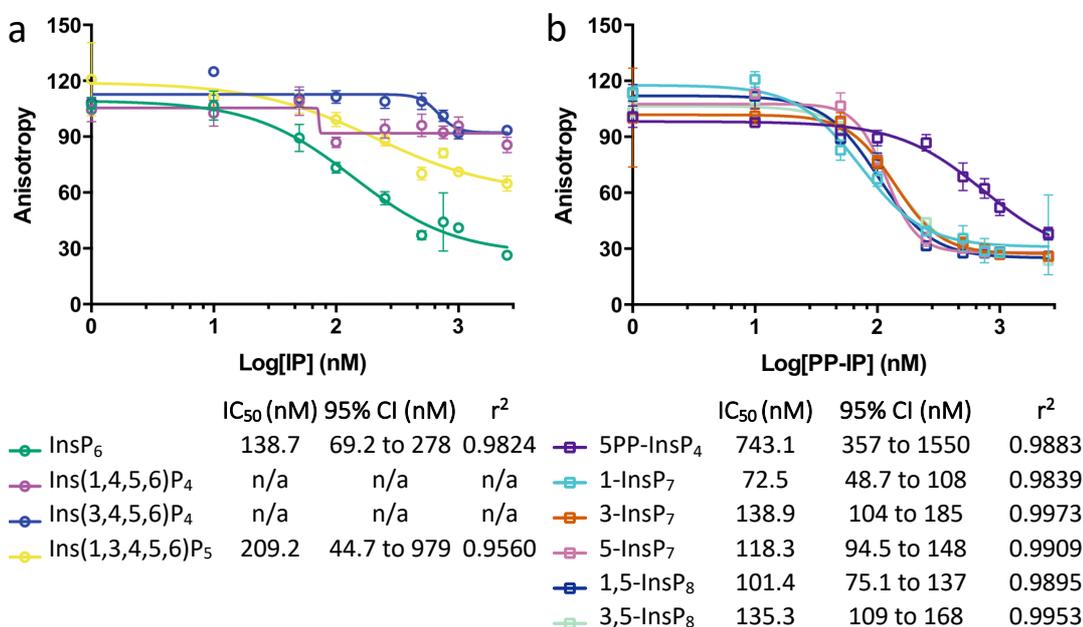


Figure 4.9 Displacement of 500 nM AtSPX1 binding 2 nM FAM-GC 40mer with titrations of (a) InsPs and (b) PP-InsPs monitored by FP. Variable slope curves generated by Prism v6.0 (Graphpad) showing the mean \pm SD of fluorescence polarization data. These binding assays were performed in 20 mM HEPES pH 6.5, 1 mM MgCl₂, 100 mM NaCl and incubated for 1 h before plate readings at 25 °C.

4.3.4 Structure predictions of SPX proteins

SPX protein tertiary structures were modelled with Phyre2 (Kelley *et al.*, 2015), and used in combination with multiple sequence alignments of solved SPX protein structures published on the RCSB PDB. This highlighted both highly conserved residues and potential residues responsible for InsP and PP-InsP binding. Unfortunately, validation of models via solving the crystal structures of SPX containing proteins of interest were not possible within the time scale of this project. Crystal trials with AtSPX1 in the conditions tested did not yield suitable crystals and without access to high throughput screening robots, this part of the project could not be taken any further.

Consequently, an AtSPX1 model (75% of residues modelled at >90% confidence) was generated using the CtVtc4 PDB structure (PDB 5IJP) as a template, as selected by Phyre2 running in “intensive” mode. When aligned in PyMol, the SPX domains from AtSPX1 and CtVtc4 are very similar [Figure 4.10; RMSD = 0.845 (113 to 113 atoms)], with the exception of the C-terminal end of AtSPX1, which is largely disordered loops connecting three short helices. The AtSPX1 model overlaid with 5IJP (Figure 4.10b) allows visualisation of the highly conserved InsP binding residues, where modelling using this structure as a template has placed these residues in very similar positions physically. Also using this alignment, the InsP₆ ligand from 5IJP can be visualised in the AtSPX1 model (Figure 4.10c), demonstrating that all expected co-ordinating residues (Figure 4.10d) are in proximity to the InsP₆. This alignment also allows us to consider the likely position of InsP binding with regards to phosphate numbering on the inositol ring. Surprisingly, since this is a modelled structure with the InsP₆ superimposed by alignment with 5IJP, there are no clashes with peptide, although some bond distances are considered too close at 2.3 Å. The InsP is exposed on the surface of AtSPX1, with only the P2 binding into a pocket formed by the N terminal section of the α 1 helix, the α 3 and the α 4 helices. The P3 position would seem restricted by the α 1 helix, and it is at this position that the

contacts with the residues on the $\alpha 1$ helix would be considered too close at 2.3 Å. The other phosphate positions appear unrestricted and fully surface accessible.

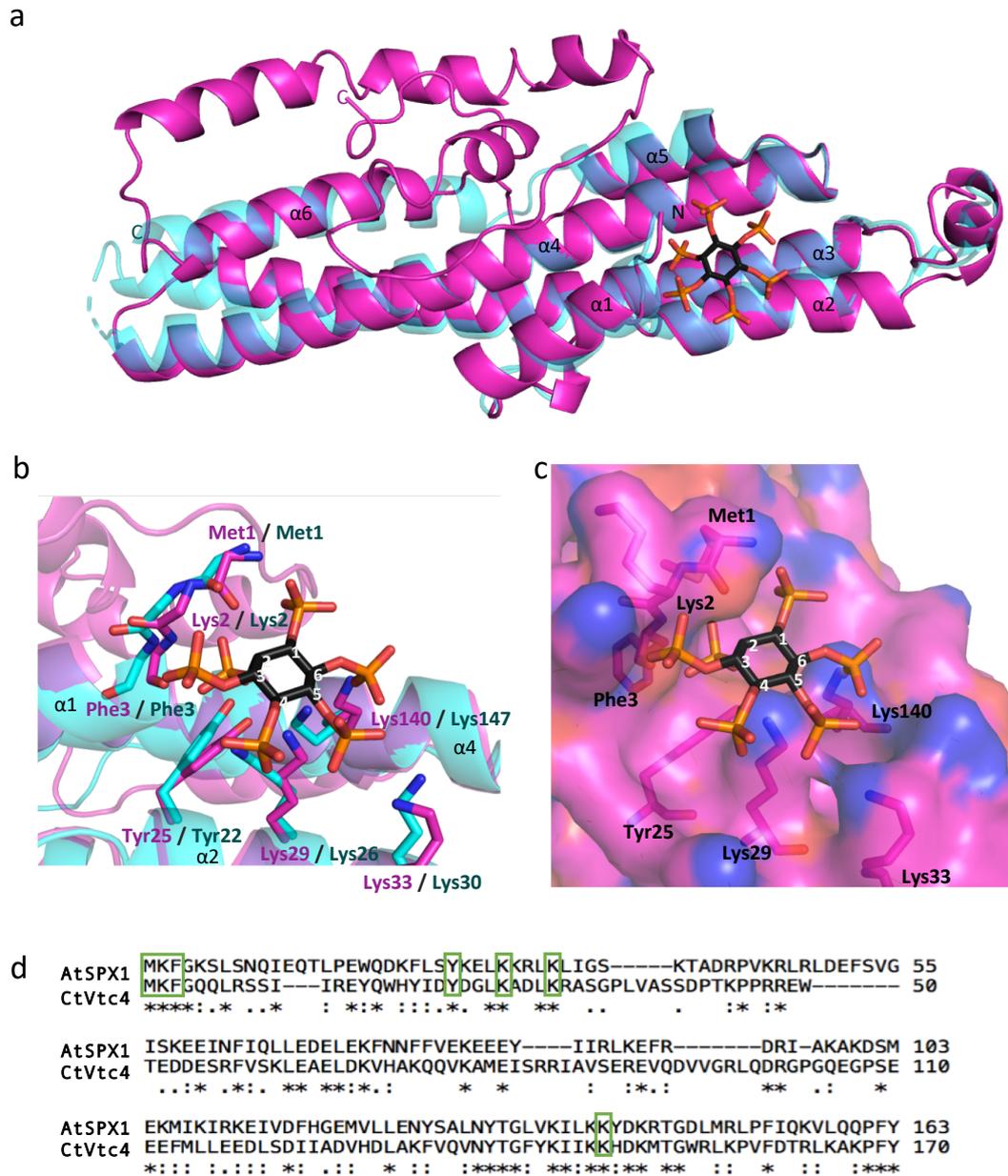


Figure 4.10 Phyre2 model of AtSPX1 (magenta) aligned with SPX^{CtVtc4} (cyan) in complex with InsP₆ (black) from PDB 5IJP structure in PyMOL. (a) Proteins structures are represented by cartoon and surface, 0.6 transparency, with conserved helices and N/C-term labelled. (b) ‘Zoom in’ of InsP₆ and contacting residues (sticks). (c) Surface of AtSPX1 (0.3 transparency) showing pocket co-ordinating axial P-2 of InsP₆, and InsP₆ contacting residues (sticks) (d) Clustal Omega alignment of amino acid sequences, highlighting corresponding InsP₆-interacting residues with green boxes.

The DP-server was used to predict DNA binding residues from the AtSPX1 amino acid sequence (Hwang *et al.*, 2007). This web server predicted fifteen sites, including several larger areas of 5-10 residues (resi 2-6, resi 142-148, resi 210-214 and resi 222-230) and individual residues/two adjacent residues (Figure 4.11). Within the putative DNA binding sites identified, three residues Lys2, Phe3 and Lys33 are in common with InsP ligand binding, suggested through conservation (Figure 4.10). However, most residues of predicted DNA binding sites are C-terminal of the SPX sensor domain.

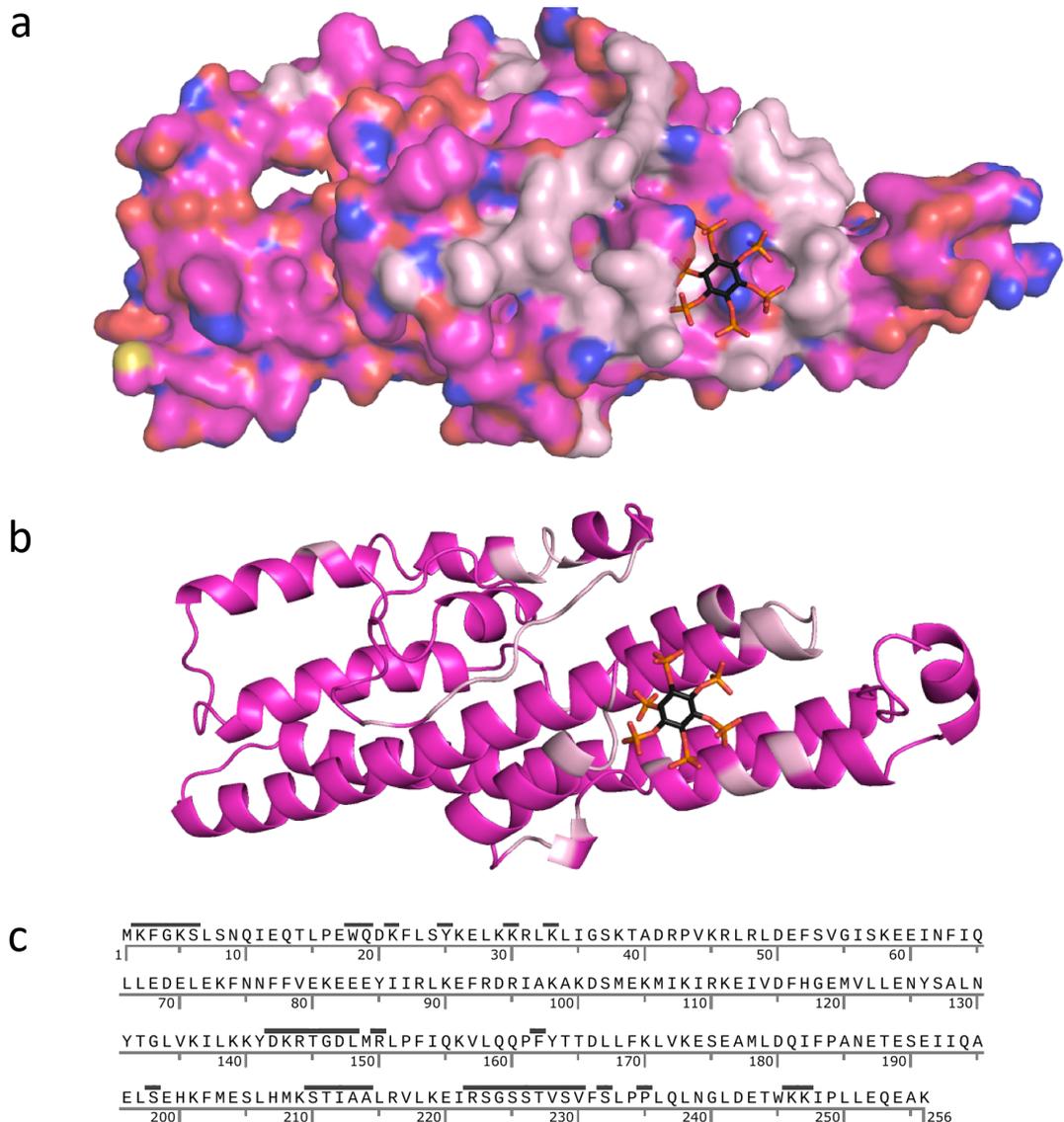


Figure 4.11 Phyre2 model of AtSPX1 (magenta) with putative DNA binding residues highlighted and InsP₆ (black) from PDB 5IJP structure in PyMOL, AtSPX1 structure is represented by (a) surface and (b) cartoon, with DNA binding sites shown in pale pink. (c) the amino acid sequence of AtSPX1 with all putative DNA binding residues by DP server highlighted by black lines (above residues).

4.4 Discussion

Within this study, a purification protocol to obtain untagged AtSPX1 is reported. Despite AtSPX1 being the simplest of SPX proteins, its purification has proved difficult, precluding others from crystallographic interrogation and biophysical study. By adopting a fluorescence polarization (FP) approach that has been particularly effective in study of Arabidopsis inositol pentakisphosphate 2-kinase, AtIPK1 (Whitfield *et al.*, 2018) and the human inositide 5-phosphatase SHIP2 (Whitfield *et al.*, 2021) ligand binding of InsPs and PP-InsPs to AtSPX1 has been achieved (as described in section 3.3.4).

A direct comparison of low and high A260/280 ratio AtSPX1 purification products revealed that AtSPX1 lacking significant nucleotide contaminant interacts with DNA *in vitro*. The AtSPX1-DNA interaction was shown to displace and be displaced by InsP ligand binding. In the absence of a protein or DNA binding partner, AtSPX1 was shown to have higher binding affinity to more phosphorylated InsP ligands, where $\text{InsP}_4 < \text{InsP}_6 < \text{InsP}_{7/8}$ and no discrimination was detected between positioning of P groups around the inositol ring. However, greater specificity was observed in displacement assays of the AtSPX1-DNA interaction, where InsPs with P groups on the C1 and C5 position were preferred over the C3 position.

4.4.1 Production of Arabidopsis SPX containing protein

Quantitative ligand binding assays and crystallographic investigations rely upon high yields (typically >5 mg, dependent on protein size) and high purity (at least 90%) of a protein of interest (McPherson *et al.*, 1996). Amongst characterization studies of Arabidopsis SPX domain proteins, with the exception of an AtSPX1 fusion with N-terminal Glutathione-S-transferase (GST), there has been little success in producing pure protein for *in vitro* studies (Wild *et al.*, 2016; Qi *et al.*, 2017; Dong *et al.*, 2019; Ried *et al.*, 2021). Within this study, whilst obtaining untagged AtSPX1 by systematic modification of a purification protocol, we failed to produce AtPHT5;1 and AtSPX3 for characterization. Individual proteins have unique requirements for optimal expression and purification, which can only be

determined by empirical data collation through trials of variant construct design, expression systems, induction methods and purification steps (Wingfield, 2016). As such, production can be the most time-consuming aspect of protein investigations.

4.4.1.1 Expression of SPX domain containing proteins

The pOPINF constructs containing *His-AtSPX1* and *His-AtSPX3* successfully overexpressed in ArcticExpress™ and Rosetta2 strains (Table 4.6). However, no expression of *His-AtPHT5;l* was detected in select *E. coli* strains (ArcticExpress™, Rosetta(DE3) or SHuffle®) under the induction conditions tested. Variations of induction conditions, such as lower IPTG concentrations and lower temperatures, were not exhausted within the systems available to us and more time could be spent to attempt protein production. On the other hand, *AtPHT5;l* is a membrane protein, which may account for some difficulties encountered in heterologous gene expression (Wagner *et al.*, 2006; Bernaudat *et al.*, 2011).

Heterologous membrane protein overexpression may be toxic to host cells through competition with native membrane protein production (Wagner *et al.*, 2006; Zoonens and Miroux, 2010). Toxic protein accumulation can cause cell death or growth defects in expression cultures; we observed neither, however toxic proteins can also lead to loss of recombinant plasmid especially when leaky expression occurs which is common in T7/*lac* promoter systems (Mertens *et al.*, 1995). Strategies to produce membrane proteins commonly involve using more tightly regulated promoter-host expression systems, such as using the tunable Lemo21(DE3) expression strain to control T7 RNA polymerase expression (Wagner *et al.*, 2008).

There are several other potential bottlenecks to producing functional *AtPHT5;l* for investigation including correct folding and targeting to host membrane systems for extraction (Wagner *et al.*, 2006). To assist correct folding and allow post-translational modifications, eukaryotic hosts may be beneficial, where *Saccharomyces cerevisiae* and *Pichia pastoris* are widely used (Karbalaei *et al.*,

2020). Another approach is to generate truncated proteins, avoiding the transmembrane domain (Wingfield, 2015; Wang *et al.*, 2021). In this study, single constructs of truncated SPX domain containing proteins were cloned into *P. pastoris*, using the pGAPZ α A expression vector. This led to failure in detection of protein by SDS-PAGE in small scale expression trials; indicating larger expression cultures would be needed to progress the study. As a more complicated expression system, fine tuning and optimization will be more time consuming (Karbalaee *et al.*, 2020).

4.4.1.2 Purification of SPX proteins

Although the yield of soluble vs. insoluble proteins was low for AtSPX1 and AtSPX3 (Appendix Figure A12.1), enough soluble protein was produced to attempt immobile metal ion affinity chromatography (IMAC) purification through scaling up the volume of expression cultures. During the first step of IMAC purification, His-AtSPX3 was largely lost in the flowthrough (Appendix I3), possibly due to low accessibility of the polyhistidine tag (NEB expression manual). However, this is not consistent with the predicted structure of AtSPX3, which shows the N-term domain is highly similar to AtSPX1, and so we would not expect this to largely differ in column binding affinity. Perhaps suggesting AtSPX3 was folded incorrectly. Contrarily, His-AtSPX1 was clearly detected in the eluate from UV trace and SDS-PAGE, allowing further purification.

Throughout the purification process of AtSPX1, high A260/A280 readings were observed, suggesting nucleotide contamination. In ideal samples of pure protein A260/A280 readings of 0.67 can be expected (Thermo Scientific). Higher ratios can indicate contamination of nucleotides such as DNA or RNA, and in other cases can be resultant of buffer additives such as DTT, which absorbs strongly at 283 nm in its oxidized form and changes to absorbance below 260 nm when reduced (Wilfinger *et al.*, 1997; Seo *et al.*, 2013). Since the A260/A280 ratio was reduced upon the addition of a high salt wash, whilst DTT concentration in the purification buffers were maintained at 2-5 mM, it appears that AtSPX1 obtained from *E. coli* is bound to nucleotide contaminants.

There are several reports of the instability of AtSPX1 (Qi *et al.*, 2017; Dong *et al.*, 2019). In our investigation, we observed AtSPX1 precipitation at protein concentrations over 1 mg/mL, which may be due to aggregation. We found that high salt concentration in the final buffer was required to prevent protein precipitation, after elution from the heparin column. Static light scattering experiments with GST-AtSPX1 revealed that the protein is more stable in high pH buffers containing glycerol and that additional salt did not improve stability (Qi, 2016). It would be interesting to monitor the stability of AtSPX1 obtained within this investigation and this information would further validate the functional studies. The tendency for proteins to form aggregates often results in failure at crystallization attempts (Zulauf and D'Arcy, 1992), which may explain why we did not obtain AtSPX1 crystals.

4.4.2 AtSPX1 InsP and PP-InsP ligand binding

In agreement with contemporary SPX protein-InsP ligand binding data, this study shows that AtSPX1 binds InsP₈ and InsP₇ ligands more tightly than InsP₆ (Qi *et al.*, 2017, Wild *et al.*, 2016). However, low discrimination between InsP species was suggested by small but not statistically significant differences between IC₅₀ values. This perhaps is unsurprising, as the SPX N-terminal backbone peptide co-ordinating InsP ligand binding has been described as highly plastic, possibly to accommodate several InsP ligands (Wild *et al.*, 2016). Lower InsPs in contrast performed badly at displacement of 2-FAM-InsP₅, none of which contained an axial P in the C2 position. As shown in the AtSPX1 Phyre2 model in alignment with InsP₆-bound CtVtc4 SPX protein, the axial P is possibly co-ordinated into the InsP binding pocket fixing the orientation of the molecule. From this, we note that the FAM-tag of 2-FAM-InsP₅ may obstruct specific AtSPX1-ligand binding and lower InsP probe binding affinity. It may be interesting to see if the same patterns of AtSPX1 InsP probe displacement were observed with the 5-FAM-InsP₅ probe. To validate hypothesized InsP binding residues, site-directed mutagenesis or solving the AtSPX1 structure with InsP ligand would be valuable experiments for future work.

Although high-throughput FP can produce a picture of multiple AtSPX1 ligand interactions and give clues to InsP discrimination with relative ease, observations noted from *in vitro* assays can have little relevance *in planta*. For example, despite small distinction between 1,5-InsP₈ and 5-InsP₇ AtSPX1 binding (IC₅₀ of approx. 56 and 50 nM respectively) *in vitro*, only 1,5-InsP₈ and not 5-InsP₇ could restore AtSPX1-AtPHR1 interaction in co-immunoprecipitation experiments in transgenic Arabidopsis seedlings (Dong *et al.*, 2019). Of course, it may be the case that InsP ligand binding affects interactions between AtSPX1 and binding partners other than AtPHR1, the only AtSPX1 binding partner which has so far been described. The possibility of other protein binding partners of AtSPX1 have been speculated over, as comparative transcriptomic experiments with the *spx1 spx2* double mutant against wildtype have shown differences in AtPHR1-independent gene regulation in Pi replenished conditions (Puga *et al.*, 2014). It has been suggested that SPX protein binding partners may change the discrimination of InsP/PP-InsP ligands, as the forming complex provides an extra surface for P interaction (Wild *et al.*, 2016). Therefore, future work may include investigating differential InsP ligand binding in complex with SPX protein binding partners.

It is also worth considering the relative cellular abundance of potential competing InsP and PP-InsP ligands. Measurements of the relative concentrations of different InsP and PP-InsP ligands in plant tissues have only recently been published (Riemer *et al.*, 2021) following the recent report of the sensitivity of capillary electrophoresis mass spectrometry (CE-MS) for detection of PP-InsPs (Qiu *et al.*, 2020). The method relies on preconcentration of PP-InsPs and InsPs from cellular extracts on TiO₂, and spiking of extracts with complementary isotope (¹³C)-labelled standards for each InsP or PP-InsP species (Qiu *et al.*, 2020). Even so, multiple InsP₇ isomers have now been detected in plant tissues (Riemer *et al.*, 2021), but they, like InsP₈ are almost collectively responsive to phosphate resupply after wholesale depletion of phosphate. The mM level of phosphate resupplied to long-term Pi starved plants grown in

hydroponics is likely to have changed many P-containing metabolites, beyond PP-InsPs and InsPs. The response to phosphate resupply may be as much an adaptive response as any signalling process. Moreover, the highly reversible conversion of Ins(3,4,5,6)P₄ to 5-InsP₇ mediated by ITPK1 and IPK1, dependent on nucleotide, make it highly likely that the favoured 1,5-InsP₈ ligand of SPX (Ried *et al.*, 2021) is intimately connected metabolically to lower InsPs as suggested (Whitfield *et al.*, 2020; Qiu *et al.*, 2021). The methods elaborated in this chapter would enable interrogation of InsP and PP-InsP competition under prevailing nucleotide conditions.

4.4.3 AtSPX1 interaction with DNA and InsPs *in vitro*

This study provides the first evidence of AtSPX1 exhibiting DNA binding activity. Attempts of using F-EMSA to consolidate DNA-binding activity seen in FP were inconclusive as AtSPX1 failed to enter the NATIVE gel matrix potentially blocking DNA migration through the gel. In direct conflict with the apparent AtSPX1-PIBS binding observed here (Figure 4.8), Qi *et al.* (2017) found that GST-AtSPX1 does not interact with the PIBS element. This discrepancy may be explained because AtSPX1-DNA binding activity was only observed in the present study after reduction of nucleotide contamination by additional purification steps. Contaminant DNA may for example competitively inhibit the AtSPX1-PIBS interaction, although the identity of the contaminant nucleotide species was not known.

There are currently no known DNA-binding SPX domain-containing proteins for comparison of equilibrium constants. The equilibrium constants obtained from AtSPX1 interaction with different oligomers fall within the lower affinity binding reported of plant transcription factors, which vary from pM to μ M range (Hao *et al.*, 1998). Whilst AtSPX1 appeared to have preference for the PIBS element over randomized probes (K_d of $\sim 0.6 \mu$ M compared to $\sim 1 \mu$ M for PIBSr), oligomer library screening experiments such as protein binding microarrays (Berger and Bulyk, 2009) or SELEX (Chai *et al.*, 2011) would be necessary to obtain consensus DNA binding sites. It is also possible that AtSPX1 binds to DNA exclusively non-

specifically (Thompson and Woodbury, 2001). Future work may look at validating the AtSPX1-DNA interaction *in planta* and using structural studies to understand the mode of DNA binding.

In this study, the AtSPX1-DNA interaction was displaced by InsPs and preferentially by PP-InsP ligands, where the most effective displacement was shown by InsP ligands with P groups on C1 and C5. A dataset using 4x PIBS as a probe confirmed the same pattern of displacement by InsPs and PP-InsPs described above. A sample of rac-1,5 InsP₈ (racemic InsP₈) became available from Megan Shipton (Potter lab, University of Oxford) in the late drafting stage of this thesis. Interestingly, in this data set chiral 1,5-InsP₈ was compared to rac-1,5 InsP₈ (equimolar 1,5-InsP₈ and 3,5-InsP₈) and the latter was not as effective mol per mol at displacing the DNA probe. This suggests again that a PP on the C3 position is less favourable for displacement of DNA compared to a PP on the C1.

The electrostatic interactions that facilitate the AtSPX1-DNA binding may also explain the displacement by highly anionic InsPs and PP-InsPs. It has been suggested that conformational changes of SPX domains occur upon InsP ligand binding (Wild et al., 2016), which may affect SPX protein binding partners and in our case DNA binding. The reciprocal experiment where 2-FAM-InsP₅ was displaced by DNA titrations was also demonstrated (Appendix 18). Several putative DNA binding residues of AtSPX1 were predicted through the DP server (Hwang *et al.*, 2007), including N-terminal residues that form part of the nuclear localization peptide (amino acids 30-43), and those that potentially interact with InsP ligands, namely Lys2 and Phe3 (Duan *et al.*, 2008). DNA inhibition of InsP binding may therefore be competitive, such that it occludes the InsP binding site. However, potential DNA binding residues were noted C-terminally of the defined SPX domain, so we cannot rule out allosteric inhibition. The mode of inhibition could be investigated by obtaining an IC₅₀ ratio from two different ligand concentrations (Wei *et al.*, 2007). Future work may also include targeting some of the putative DNA binding residues by site-directed mutagenesis and exploring any differences in DNA binding activity.

4.4.4 Implications of AtSPX1 DNA binding activity

A possible model of DNA binding activity in the role of PSR could be that under Pi starvation marked by low availability 1,5-InsP₈ (Dong *et al.*, 2019), but also InsP_{7s} (Riemer *et al.*, 2021), AtSPX1 binds to DNA non-sequence specifically but preferentially near the PIBS element. This may serve to promote the AtPHR1-PIBS by indirect co-operativity, following the theory that higher abundancies of DNA binding proteins improve the probability of transcription factors locating target sequences, by competing with nucleosome binding over nucleotide availability (Morgunova and Taipale, 2017). Conversely, low affinity binding of AtSPX1-PIBS may directly and weakly regulate against specific AtPHR1-PIBS target site binding (Zabet and Adryan, 2013). Or it could be a mechanism to localize the AtSPX1 in proximity to AtPHR1. In either case, AtPHR1 is able to activate PSI genes for PSR. When Pi is replenished and 1,5-InsP₈ is available, the AtSPX1 DNA binding activity is displaced and AtSPX1-AtPHR1 interaction is promoted, inhibiting AtPHR1-PIBS and negatively regulating PSR (Dong *et al.*, 2019). A major problem of this model is that within this investigation, many InsP ligands were capable of inhibiting the AtSPX1-DNA interaction, including the highly abundant InsP₆ and high affinity 1/5-InsP₇ ligands (IC₅₀ ~70-120 nM), both of which have been detected in similar levels under Pi deplete or replete conditions in Arabidopsis seedlings (Dong *et al.*, 2019) and, most convincingly, under Pi resupply after long-term starvation (Riemer *et al.*, 2021). Therefore, we are brought back to the question of the tissue and subcellular organization and availability of InsPs/PP-InsPs. Specifically, what is the availability of higher InsP ligands in the nucleus and how does that compare with available DNA, and free nucleotides, perhaps, to interact with?

The subcellular abundancies of different InsPs/PP-InsPs are not known, however the enzymes responsible for PP-InsP production have nuclear and cytoplasmic localization (Xia, 2003; Kuo *et al.*, 2014; Adepoju *et al.*, 2019), which suggests pools of these molecules (InsP₆, InsP₇ and InsP₈) are available in the nucleus. Within the nuclear environment, the high density of DNA likely means that non-

specific DNA binding would happen *in vivo*, as we observed *in vitro*. Although DNA concentration has not been quantified in plant nuclei, accessible DNA between nucleosomes (on average 56 bp) has been estimated in humans at approx. 0.5 mM (Valouev *et al.*, 2011), which would be sufficient for non-specific DNA binding interactions occurring at micromolar binding affinities. However, accessible stretches of DNA are competed for by other DNA binding proteins (to regulate transcription, or other nuclear activity) and may be densely covered by/in competition with such proteins (Klemm, Shipony and Greenleaf, 2019).

4.5 Conclusion

In this chapter, *in vitro* binding of InsP ligand to AtSPX1 has been described and evidence presented of DNA binding activity regulated by InsP ligands. Despite limitations of *in vitro* studies, they provide biochemical evidence to assess the plausibility of possible *in vivo* interactions. Further work to understand SPX mediation of Pi homeostasis may involve identifying and characterizing SPX protein binding partners alongside InsP ligand binding. However, studies that reveal the abundancies of InsPs and subcellular compartmentalization remain a crucial milestone to achieving an understanding of plant Pi homeostasis and PSR.

5 General conclusions and future work

Metabolism of inositol phosphates (InsPs) and their binding to cognate partners has far reaching importance for plants, in capacities such as basal pathogen resistance (Poon, 2019), phosphate homeostasis (Ried *et al.*, 2021) and other stress tolerances (Jia *et al.*, 2019). Although specific InsPs have been implicated in various cellular roles, assignment of function is challenging due to the technical difficulties of InsP detection and identification, their rapid turnover in cells, and functional redundancy of enzymes interconverting them or of the inositol phosphates themselves. Similarly, the roles of InsP metabolizing enzymes are not always clear, nor is their activity in isolation of the InsPs they bind. This work presents evidence of additional function of InsP-binding proteins that extends beyond their interaction with inositol phosphates, either as substrates or cognate ligands.

Novel DNA binding activity of two plant InsP-binding proteins, AtIPK1 and AtSPX1, was explored using biochemical and biophysical assays. AtIPK1 is crucial for InsP₆ synthesis, important for Pi homeostasis (Kuo *et al.*, 2018), basal plant defence (Murphy *et al.*, 2008), seedling viability (Kuo *et al.*, 2014), messenger RNA export (Lee *et al.*, 2015) and has partial nuclear localization (Kuo *et al.*, 2014), yet a nuclear role (independent of kinase activity of InsP₅ (2'OH) substrate) for AtIPK1 has not been defined. Similarly, with localization in the nucleus (Duan *et al.*, 2008), AtSPX1 is a receptor for 1,5-InsP₈ which modulates AtPHR1 in transcription regulation of PSR genes (Puga *et al.*, 2014; Wild *et al.*, 2016; Dong *et al.*, 2019), without known enzymic activity. Previous characterization of AtSPX1 has been interpreted in context of its interaction with its protein binding partner AtPHR1, evidenced in pull-down assays but limited in the context of analysis of ligand binding (Wild *et al.*, 2016; Ried *et al.*, 2021). This in part has been due to difficulties in production of soluble SPX protein, for which we had similar problems when attempting to produce soluble AtSPX3 and AtPHT5;1. In this work, high affinity non-sequence specific DNA-binding (K_d of

~0.4 and ~0.6-2 μM) was observed *in vitro* for both AtIPK1 and AtSPX1, pointing to biological relevance *in planta*.

Fluorescence polarization was utilized to interrogate ligand binding affinities for AtIPK1 and AtSPX1, specifically using inhibitor screens with InsPs and other ligands. High-throughput screening for inhibitors of AtIPK1 (and the inositol phosphate 5-phosphatase SHIP2) has been described recently with the 2-FAM-InsP₅ probe used in this work (Whitfield *et al.*, 2018, 2021). Extension of this approach to AtSPX1 provides an easily accessible methodology that could find use with the expanded SPX family of proteins in plants. The method uses minimal amounts of protein, is rapid and could easily be automated for high-throughput screening of compound libraries. Following demonstration that AtSPX1 binds 2-FAM-InsP₅, it was shown that regiospecific distribution of phosphate substituents of different displacing InsP ligands was, in the absence of DNA, less important than overall anionic charge to the interaction of inositol phosphate with AtSPX1.

Protein-DNA interactions were also studied by the polarization approach, here monitoring polarization changes of FAM-tagged oligonucleotides upon addition of different InsP ligands. The AtIPK1-DNA complex was displaced by additions of InsP₅ or InsP₆ (required at a higher dosage threshold for effect) and the active site inhibitor purpurogallin, whereas nucleotides (such as ATP and ADP) did not interfere with AtIPK1 DNA binding activity. Thus, revealing that DNA binding of AtIPK1 is inhibited by InsP substrate or product binding at the active site.

In the case of AtSPX1, whose purification is enabled by the removal of nucleotides, an extended library of InsP and PP-InsP ligands were screened. From those tested, lower InsPs (including InsP₅) were poor inhibitors of AtSPX1-DNA binding and the highly phosphorylated ligands were more effective. Amongst this screen, regiospecific discrimination was observed amongst the tested InsP_{7S}, which exhibit the same anionic charge yet clear preference for pyrophosphate on D1 >> D5 > D3 carbons respectively. This suggests that DNA binding activity of

AtSPX1 is regulated by multiple InsP and PP-InsP ligands. Inversely, it was shown that DNA binding to AtIPK1 or AtSPX1 prevents InsP ligand binding, here 2-FAM-InsP₅ was displaced with untagged DNA titrations. These studies raise the possibility, for both AtIPK1 and AtSPX1, that InsP-binding specificities are modulated by DNA binding activity, and that DNA binding activity itself is negatively regulated by InsP binding and availability.

Extension of the polarization approach to DNA sequence binding preferences failed to identify consensus target DNA sequences, but nevertheless confirmed non sequence specific binding. These interactions are highly electrostatic and mostly occur through protein interactions with the ribose-phosphate DNA backbone (von Hippel and Berg, 1986; Boyer, 2002). Non-specific DNA binding proteins may display high affinity binding to a wide array of DNA sequences, and many have roles including DNA packaging or maintenance (Agback *et al.*, 1998). While non-specific interactions can be characterized in the same manner to sequence specific work, the generality of the interaction may hinder definition of a nuclear role. This often requires study of the mechanism of DNA binding, and structure-function studies can be a helpful approach, for example, for confirmation of DNA binding motifs (which will be further discussed below). Nevertheless, use of FAM-tagged oligonucleotide may allow high screening throughput of promoter elements to determine a consensus binding site, and therefore begin to assign specific processes that may be regulated by AtIPK1 or AtSPX1 DNA binding activity.

The non-specific DNA binding observed may simply represent a “parking spot” for AtIPK1 and AtSPX1, such as that modelled for non-specific DNA binding ligands (Flyvberg *et al.*, 2006), which is negatively regulated by the availability of InsPs. When non-specific DNA binding occurs, or when these proteins are “parked”, the efficiency of sequence-specific DNA binding protein activity may be reduced by the occlusion of target binding sites (Flyvberg *et al.*, 2006). For example, AtIPK1 or AtSPX1 may bind available DNA at, or close to, sites where transcription factors and other DNA binding proteins (DBPs) compete for gene expression regulation, DNA modification or DNA packaging. As proteins of

different sizes (~50 kDa for AtIPK1 and ~30 kDa for AtSPX1) it is possible the extent of DNA occluded by non-specific interactions may also differ. It would be interesting to determine the number of bp occluded by these protein-DNA interactions, by DNA footprinting experiments or solving the structures of protein-DNA complexes. In the presence of specific InsP or PP-InsP ligands, for example, when AtIPK1 binds its physiological substrate InsP₅ or InsP₆ product, and when PP-InsPs (1-/5-InsP₇ and 1,5-InsP₈) interact with AtSPX1, DNA binding activity is inhibited. The non-specific DNA binders are removed from their “parking” space and other DBPs can now bind at the free site; to regulate various parts of DNA metabolism. This is a simplified model and it is important to recognise that all protein-DNA interactions are transitory (Flyvberg *et al.*, 2006) and that non-specific DNA binding interactions are often weaker than sequence-specific binding (Boyer, 2002), therefore “parking” is only temporary even in the absence of inhibitor ligands.

AtSPX1 appeared to show preference for the PIBS element compared to random oligonucleotide sequences of the same length *in vitro*. This preference, or higher binding affinity for PIBS may increase the tendency of AtSPX1 to locate the PIBS site, the consensus binding site for the AtPHR1 transcription factor, and therefore increase AtSPX1 proximity to AtPHR1 (Figure 5.1). If DNA-interacting proteins (such as IPK1 or SPX1) bind transiently to available stretches of DNA in an overcrowded nucleus, competing proteins with higher binding affinity will be more likely to occupy consensus sites at any one moment. Therefore, we would expect that AtPHR1 dimer sequence-specific binding would outcompete AtSPX1-DNA binding at the PIBS site and might activate expression of a subset of PSR genes. Indeed, this work implies that the PIBS element is not necessarily, or exclusively, a consensus binding site of AtSPX1. The expression of *AtPHR1* does not appear to be Pi-dependent (Rubio *et al.*, 2001; Zhou *et al.* 2008) but has been shown to be positively regulated by auxin (Huang *et al.*, 2018); whereas *AtSPX1* expression has been suggested to be under the control of the PIBS promoter (Puga *et al.*, 2014) and it has recently been shown that MtSPX1 has the PIBS cis-elements in its promotor region (Wang *et al.*, 2021). Therefore, *AtSPX1* may be

upregulated by AtPHR1-mediated PSI gene activation changing relative protein abundancies. When AtSPX1 is more abundant it will better compete for PIBS sites (assuming that AtPHR1 levels do not also increase) and act as a weak negative regulator of PSI activation, including negative autoregulation to stabilize protein levels. This model applies in conditions where Pi availability is low, or inhibitory InsP/PP-InsP ligands are not available.

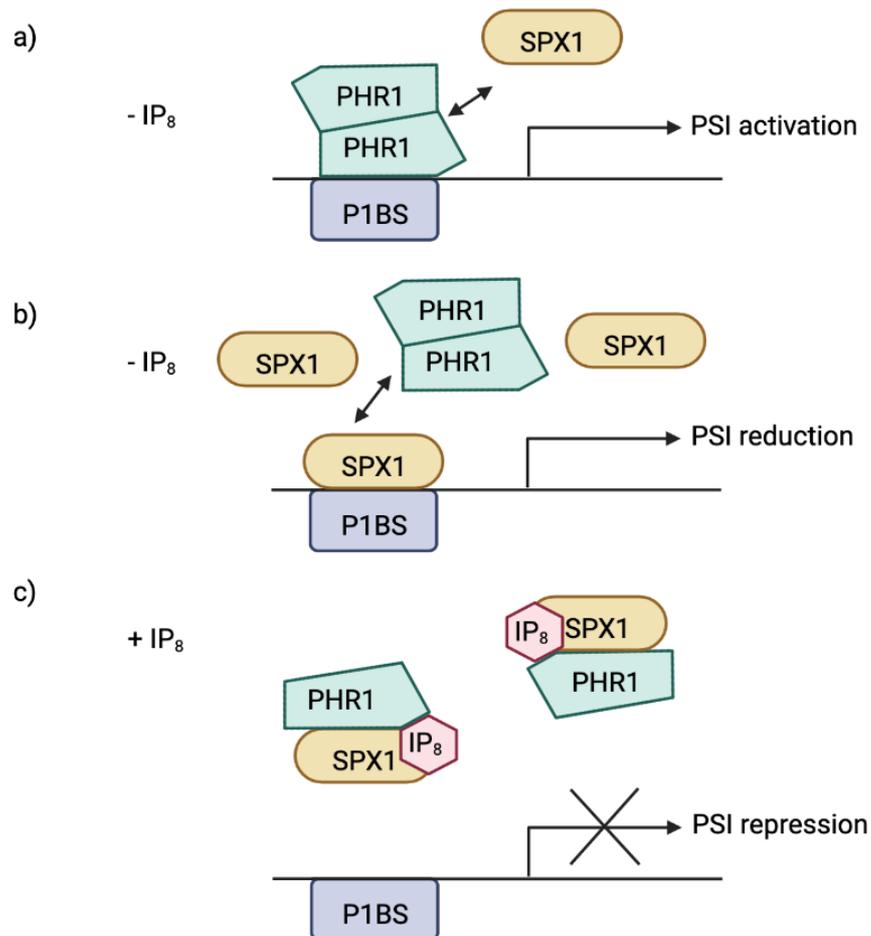


Figure 5.1 Model for AtSPX1 directly and indirectly regulating AtPHR1-mediated Pi starvation responses by binding at the PIBS site. (a) in the absence of InsP₈ ligand (phosphate deplete), AtPHR1 dimer binds to consensus PIBS site to activate target PSI genes, which is weakly regulated by a lower affinity binding of AtSPX1 at PIBS. (b) when expression levels of AtSPX1 increase more SPX1 protein is available to compete with AtPHR1 at the PIBS site and negatively regulate PSR including SPX1 expression) (c) when InsP₈ is available (phosphate replete), the complex between AtSPX1 and AtPHR1 is stabilized and PSR is repressed. Figure created with Biorender.com.

In this work it was shown that 1,5-InsP₈ was an effective inhibitor of AtSPX1-DNA binding activity, so we can speculate that if elevations of this ligand extend to the plant nucleus, AtSPX1 will preferentially bind 1,5-InsP₈, thus causing AtSPX1 to dissociate from its “Parking” spot on the DNA and bind its neighbouring physiological target, AtPHR1 (Figure 5.1c). Under phosphate replete conditions AtPHR1 transcription factor binds as a dimer at the PIBS site to activate Pi starvation responses (Rubio *et al.*, 2001) and AtPHR1-DNA binding is inhibited when AtSPX1, in the presence of 1,5-InsP₈ co-ligand, binds to AtPHR1, thus preventing dimer formation (Qi *et al.*, 2017; Ried *et al.*, 2021). While this proposed model may be supported by studies *in planta*, where some *vih* mutants, which cannot produce 1,5-InsP₈, were unable to negatively regulate Pi starvation response in Pi replete conditions (Dong *et al.*, 2019), others have questioned an exclusive role for 1,5-InsP₈ as the sole agent of repression of PSR. Kuo *et al.* (2018) and Gillaspay and coworkers (Land *et al.*, 2021) failed to observe PSR or phosphate over-accumulation in assorted *vih* (*vip*) mutants. The same cannot be said of *Atitpkl* or *Atltpkl* mutants, which without contradiction show robust phosphate over-accumulation and PSR response.

In considering interactions in Pi deplete/replete conditions, DNA and ligand binding in response to varying Pi concentrations was also studied. Upon additions of Pi, AtSPX1 appeared to bind DNA with greater affinity and, inversely, bound 2-FAM-InsP₅ ligand with lower affinity. However, on closer inspection of this work it was noted that significant lowering of the pH occurred on addition of Pi despite the presence of buffer. In the absence of Pi, at a range of pH concentrations, a similar effect was seen of AtSPX1 binding to DNA/2-FAM-InsP₅ confirming that any responses seen were as a direct result of the pH and not the Pi levels. This could suggest that ligand or DNA binding of the proteins studied here is also influenced by cell environmental cues such as cell pH, known to undergo changes under certain stresses such as anoxia (Felle, 2005). Taking into account relative levels of InsPs, cues such as pH and Pi, crosstalk of stress pathways can be considered a complicated and ever-changing network of interactions. This makes it difficult to assign unequivocal function to any

particular protein, since function may change dependent on physiological environment.

In our study, it was shown that InsP_{7s} (specifically 1 and 5-InsP₇) were also highly effective inhibitors of AtSPX1-DNA binding (equally if not more so than 1,5-InsP₈) *in vitro* and to lesser efficacy InsP₆, however physiological significance has not been demonstrated. Likewise, InsP_{7s} and InsP₆ could also displace AtPHR1-DNA interactions *in vitro* (Dong *et al.*, 2019; Qi *et al.*, 2017), though neither were deemed essential to Pi starvation regulation (Dong *et al.*, 2019). It may be worthwhile to challenge the emphasis on 1,5-InsP₈ given the observations of Kuo *et al.* (2018) and Land *et al.* (2021) and because much of the biochemistry of the VIH1/2 protein has been studied with isolated domains of the protein, not the full-length entity, nor has the reversibility of the kinase activity been considered. Here, it is important to consider relative levels of InsPs, PP-InsPs, ADP and ATP, along with binding affinities of the proteins and complexes that bind them. The abundance of 1,5-InsP₈ ligand is increased on phosphate resupply after chronic Pi starvation, but so are the levels of all other inositol phosphates (Reimer *et al.*, 2021). Without a clear picture of compartmentalization of InsP/PP-InsPs within plant cells or full biochemical description of full-length VIH1/2 protein, there remains considerable uncertainty.

The nuclear environment of eukaryotic nuclei, where DNA metabolism occurs, is enclosed by the nuclear envelope. The nuclear envelope has several nuclear pore complexes which facilitate the movement of proteins and nucleotides by active transport, as well as allowing diffusion of water-soluble molecules between the cytoplasm and nucleoplasm (Alberts *et al.*, 2002). It is possible that water-soluble InsP molecules are exchanged between the cytoplasm and nucleoplasm (Figure 5.2). At present, it has been demonstrated that the lipid-dependent and lipid-independent molecular apparatus for InsP and PP-InsP synthesis is available in plant nuclei (Dieck *et al.*, 2012), including AtIPK1. Therefore, the production and turnover of these molecules is expected to occur here, as well as the cytosol. The availability of InsP/PP-InsP-binding proteins in the nucleus is

also possible, as evidence of nuclear localization have emerged, including AtSPX1 (Puga *et al.*, 2014), the auxin receptor TIR1 (Prigge *et al.*, 2020) and COI1 (Withers *et al.*, 2012). These proteins may sequester specific InsP/PP-InsP ligands as they form complexes with target proteins. Combined with the considerations discussed above, it is clear that both AtSPX1 and AtIPK1 experience a complex cellular environment dependent on growth conditions, impossible to replicate *in vitro*, but is something that we must consider when postulating novel roles for these proteins.

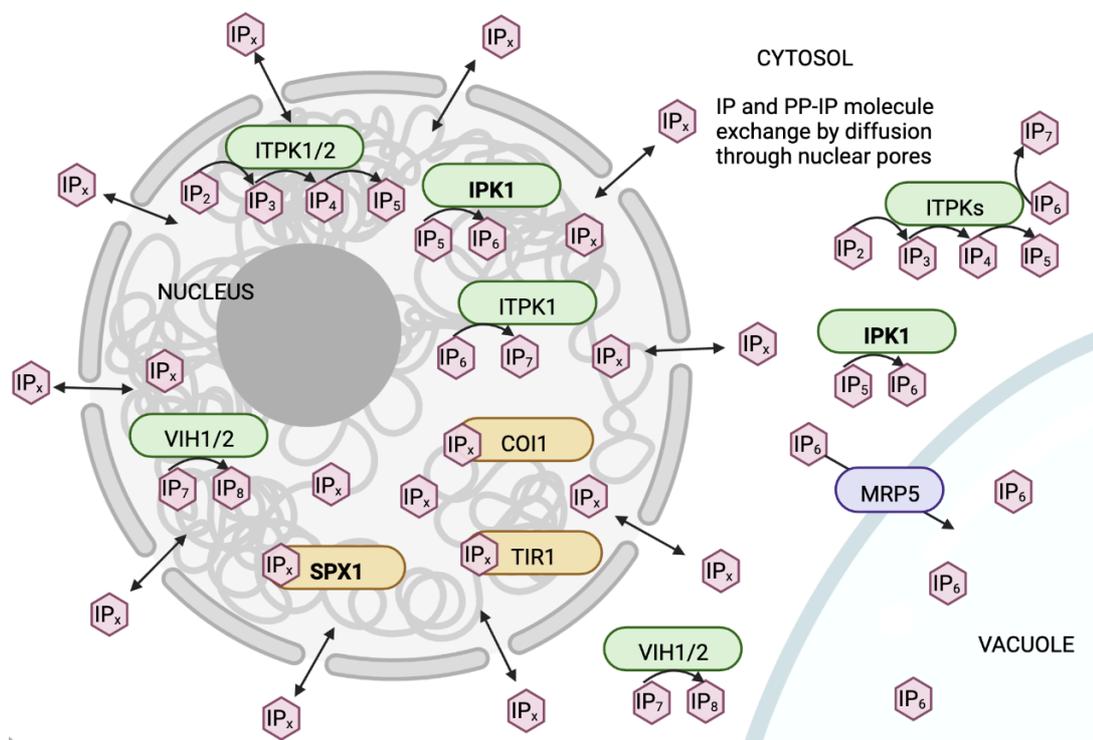


Figure 5.2 A simplified schematic of InsP metabolism and transport between organelles (nucleus, vacuole) and cytosol in a plant cell. Highlighting IPKs which have been shown to localize in the nucleus from the lipid independent pathway of InsP biosynthesis (green) and InsP binding proteins (yellow). InsPs are shown as pink hexagons. Straight arrows represent direction of movement of InsPs, curly arrows represent phosphorylation of InsPs. Figure created with Biorender.com.

Within the *in vitro* DNA binding assays applied in this study, 50 mM and 100 mM NaCl (for AtIPK1 and AtSPX1 experiments respectively) were used to obtain EC₅₀, IC₅₀ or K_d values. These salt concentrations do not mimic physiological

conditions; under typical physiological conditions, glycophytic plants (such as *Arabidopsis*) have a relatively high cytosolic concentration of K^+ ions (100-200 mM) and low concentrations of Na^+ (1-20 mM), Cl^- (5-20 mM) (Blumwald *et al.*, 2000; Geilfus, 2018) and Mg^{2+} ions (0.2-0.4 mM) (Hermans *et al.*, 2013). However, for the purpose of limiting non-specific interactions NaCl and KCl have similar ionic strengths. Therefore, AtSPX1-DNA binding assays performed with 100 mM NaCl are more reflective of physiological conditions. Additionally, 1 mM $MgCl_2$ was used in binding assays, approx. 2-5x the concentration of free Mg^{2+} that is expected from the cytosol (Hermans *et al.*, 2013). However, as well as free Mg^{2+} availability, this ion is commonly found bound with nucleotidyl phosphates (Hermans *et al.*, 2013) and as cofactor to many proteins (Pilchova *et al.*, 2017), including AtIPK1 [where it facilitates kinase activity by co-ordinating nucleotide recognition via G-loop residues (González *et al.*, 2010)]. Mg^{2+} can also stabilize the duplex DNA conformation (Owczarzy *et al.*, 2008) and may affect protein-DNA binding, wherein specific interactions are promoted and non-specific interactions are reduced (Moll *et al.*, 2002). In trials with F-EMSA and AtIPK1, the addition of 1 mM EDTA which chelates Mg^{2+} was used as a variable for optimizing binding (considering that $MgCl_2$ may destabilize DNA binding). In these experiments it was shown that Mg^{2+} did not either greatly hinder or promote the AtIPK1-DNA interaction.

Establishing whether AtIPK1 or AtSPX1 DNA binding occurs *in planta* is crucial to determine physiological relevance. DNA binding proteins serve a multitude of physiological functions within cell biology, for example in regulation of gene expression, programmed cell death and cell division. Within this work, AtIPK1-DNA binding interaction *in planta* was demonstrated using FLIM-FRET: reflected in reduced fluorescence lifetime of GFP-tagged AtIPK1. While improvements to the protocol will enable others to confirm the results of this study, extension of the approach into different genetic backgrounds and to different protein targets, including AtSPX1, promises to be an important tool. Alternative methods to explore a protein-DNA interaction *in vivo* may involve ChIP-seq, allowing identification of potential physiological targets or consensus binding sequences, to answer whether specific DNA binding occurs for either protein. DNA binders

such as transcription factors can bind specific target sequences to recruit or protect from DNA modifying enzymes such as endonucleases or polymerases. Although high throughput screening of DNA libraries (e.g., using SELEX) would be a powerful method to find DNA consensus sequences, it is costly and perhaps screening a modest-sized set of synthesized oligonucleotides taken from promoter regions of InsP metabolizing/interacting genes would be a cheaper alternative to start.

In silico modelling was adopted to gain insight into the structure-function relationships of DNA binding and substrate recognition. Site-directed mutagenesis and X-ray crystallography screening were also brought to bear in this thesis. The structure of AtIPK1 has been solved several times, with numerous structures available on the PDB at the time of writing. In contrast, the AtSPX1 structure remains unsolved (NB after this thesis had been completed, the structure of OsSPX1 was solved PDB ID 7E40, Zhou *et al.*, 2021); although homology-based modelling using SPX domain containing proteins as a reference point allowed simplistic examination for this small, single-domain protein within this work. A challenge presented to this study of both proteins is the lack of structural homology between known DNA binding proteins and AtIPK1/AtSPX1. This itself may suggest a novel mechanism of DNA binding.

Whilst working with more structural data available for AtIPK1, docking was attempted with dodecamer DNA (PDB 3UXW) as a potential ligand with HDOCK, and different sites for DNA binding were identified: the InsP binding pocket, the zinc finger motif residues, and two small tripeptide 'hooks' GRP and GKE which may act co-operatively. Within this thesis, three AtIPK1 GRP mutants were generated including a deletion of this tripeptide mutant [referred to as Del(GRP)], and DNA binding activity were examined with comparisons to wt AtIPK1. It was concluded that this tripeptide motif is not essential for non-specific DNA interactions. This does not exclude the possibility that this motif could interact with DNA sequence-specifically if AtIPK1 has a consensus binding site. However, it is likely that non-specific DNA binding is facilitated by an alternative AtIPK1 feature.

The InsP binding surface of AtIPK1, also referred to as the C-IP lobe, is a highly conserved structural feature of IPKs, as they specifically co-ordinate the Ins(1,3,4,5,6)P₅ substrate for kinase activity. Within the sequence conservation analysis run by ConSurf-DB, there are many highly conserved C-IP lobe amino acids amongst AtIPK1 homologues including critical residues for catalysis; for example, Lys168, Asp368, Lys411. If the AtIPK1 InsP binding surface is involved in DNA binding, it may follow that other IPKs from different kingdoms can bind to DNA. In contrast, the GRP/GKE motifs were not particularly conserved amongst IPK homologs and are small structural features (in comparison to the extensive C-IP or positively charged binding surface). If the GRP/GKE tripeptides are present in the amino acid sequence data of homologues it is important to consider whether they would be available on the surface of the protein to interact with DNA, therefore additional structural data would be necessary to predict DNA binding. There are several routes for further investigations: crystallization screens may be improved by targeted screening (e.g., using oligomers containing a preferred DNA binding sequence), active site AtIPK1 mutants may be examined for DNA binding activity, SAXS analyses of AtIPK1-DNA may give insight to the DNA-binding site and be used to improve modelling, and further molecular dynamics simulations may allow assessment of DNA binding physics.

In AtIPK1-DNA binding assays, catalytically active AtIPK1 (demonstrated using HPLC to follow formation of ADP by-product) was used without substrates in the binding buffer. These experiments demonstrate that the AtIPK1-DNA interaction is not dependent on the presence of InsP or nucleotide substrates/catalytic products and as discussed above, InsP binding inhibits AtIPK1 DNA binding activity. AtIPK1 catalysis is thought to occur through conformational changes upon substrate binding which stabilize the N-lobe and enable catalytic activity. In the absence of substrates, AtIPK1 has increased flexibility of the N-lobe (Baños-Sanz *et al.*, 2012; Gosein and Miller, 2013) compared to more closed conformations, as further demonstrated by the apo and nucleotide bound wt AtIPK1 structures solved in this thesis. It is possible that flexibility of the N-lobe

is important for DNA binding. This may be supported by *in silico* docking data (Appendix 10) which show that DNA binding potential on the protein surface is altered dependent on ATP bound or ternary complexes of AtIPK1 as input models in HDOCK. It would be interesting to test the hypothesis that the N-lobe stability affects DNA binding, employing the use of AtIPK1 disulfide mutants, where an artificially stabilized AtIPK1 N-lobe mutant [such as in Gosein's AtIPK1 E82C/S142C (Gosein and Miller, 2013)] is used in DNA binding assays alongside a positive control of wildtype AtIPK1 for comparison.

There are several AtIPK1 active site mutants which have been shown to be catalytically inactive, such as alanine mutants of Arg130, Lys168, Asp368, Lys411 (Gosein and Miller, 2013). However, the authors have not shown whether the mutants retain InsP₅ binding capacity. InsP₅ binding capacity may be tested by using a 2-FAM-InsP₅ probe in FP assays and screening binding with AtIPK1 mutants compared to wt AtIPK1. The Brearley lab has collaborated with Beatriz González's lab in the past to generate AtIPK1 active site mutants. In some preliminary FP experiments (not shown in this thesis) using wildtype AtIPK1 vs W129A, R130I and W129A/R130I double mutant, all four proteins could bind to a FAM-DNA. Due to limitations of how much protein was available of these mutants, it was not possible to obtain data to compare the binding affinities.

In this work, potential DNA-binding residues of AtSPX1 were also examined. However, due to lack of a crystal structure, a sequence-based predictive webserver was used for modelling ligand binding and identifying important residues. Interestingly, there was some overlap in AtSPX1 residues which were predicted to bind DNA and those predicted to bind InsP substrates, so we may speculate that DNA or InsP binding occurs at one site of AtSPX1. Future work may involve targeting these residues to test the hypothesis of their involvement in DNA binding and importance to ligand binding. In the phyre2 model of AtSPX1 (Figure 4.10) generated in this study, amino acid residues 1-3 from α 1 appear to help anchor the P-groups at positions D1, D2 and D3 of InsP₆. However, in the recently solved structure of OsSPX1 in complex with PHR2 (Zhou *et al.*,

2021), although these conserved residues were important for binding affinity of InsP₆, they were not essential whereas residues Y25/K29/K151 from α 2 helix were crucial. The InsP₆ ligand placed in the OsSPX1-PHR2 complex also appears to be in a different orientation to the InsP₆ superimposed in the AtSPX1 model shown in this work, highlighting the importance of solving the structure of AtSPX1.

Similar to AtIPK1, conformational changes of SPX domains upon InsP and PP-InsP ligand binding have been reported (Wild *et al.* 2016; Zhou *et al.*, 2021), wherein a shift of the α 1 helix has been observed. This conformational change is important for OsSPX1 to pack tightly and form a stable 1:1 complex with PHR2 preventing PHR2 dimerization (Zhou *et al.*, 2021), and thus enabling repression of PHR2-mediated PSR. It is possible that conformational changes of AtSPX1 dependent on specific InsP/PP-InsP binding can also exclude DNA binding. A potential follow-up line of investigation may be to perform NMR titrations with the full length AtSPX1 protein to monitor conformational shifts with different InsP substrates, and attempt crystallization or SAXS analysis of this protein in the absence of and with different ligands including DNA.

Although both AtIPK1 and AtSPX1 bind InsPs/PP-InsPs via clustered groups of basic amino acids, InsP receptor/binding proteins (also including TIR1 and COII) have less defined InsP binding motifs compared to IPKs; and there is little structure or sequence homology between them (Alcázar-Román, 2008). This gap may narrow as more structures of proteins that bind InsPs are solved with InsP ligand and other interacting partners e.g. HDAC3-MTA1-InsP₆ (Watson *et al.*, 2016). Nevertheless, in validation of DNA binding to AtIPK1 we have shown that the presence of an InsP binding pocket is not in itself a prerequisite to DNA binding, since AtITPK4 did not bind DNA. While this may have been anticipated, it is a powerful control. Future work may yet elucidate the mechanism of structural communication between DNA binding and InsP occupancy of the active site/ligand binding. If we can identify and verify a DNA binding site from AtIPK1 or AtSPX1 it may be possible to search for similar motifs in homologous proteins across kingdoms.

Taken together, the studies in this thesis present the possibility of additional nuclear functions for known InsP-binding proteins via interaction with DNA. This extra-activity was shown to be regulated by InsP and PP-InsP *in vitro*, from which we propose that nuclear regulation by AtIPK1 or AtSPX1 interaction with DNA may be further regulated by InsP/PP-InsP signal molecules, including physiological substrates/ligands. This adds another dimension to the complicated story of InsP mediated signaling in plants. As discussed previously, much work is needed to assign roles to these newly described characteristics for either AtIPK1 or AtSPX1; however, perhaps this work will encourage exploration of novel roles for previously assigned InsP interacting proteins.

Appendices

Appendix 1 The 2-FAM-InsP₅ ligand

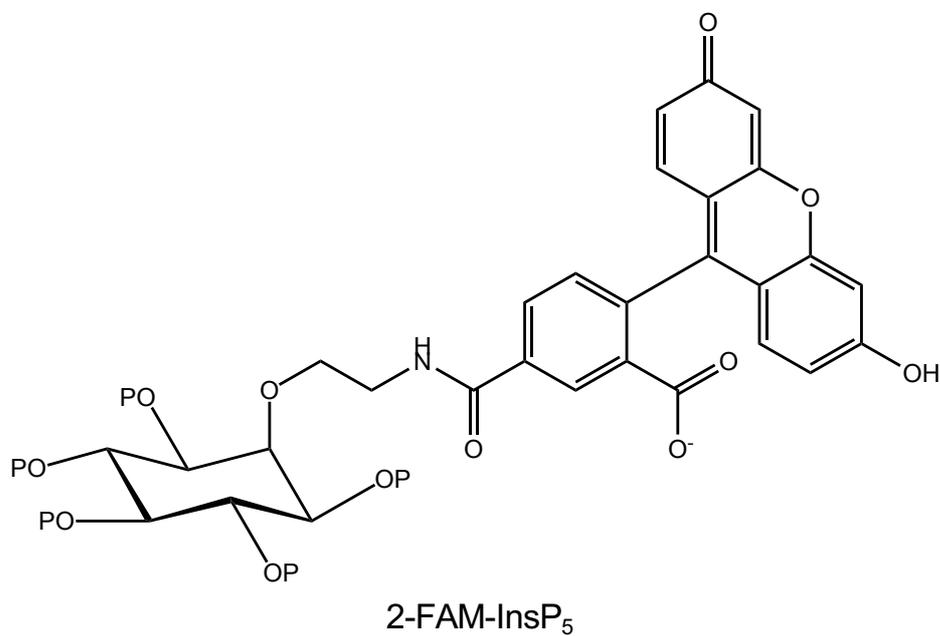


Figure A1 The structure of 2-FAM-InsP₅ used in FP binding and displacement assays. Made in ChemDraw.

Appendix 2 Purification of wt AtIPK1

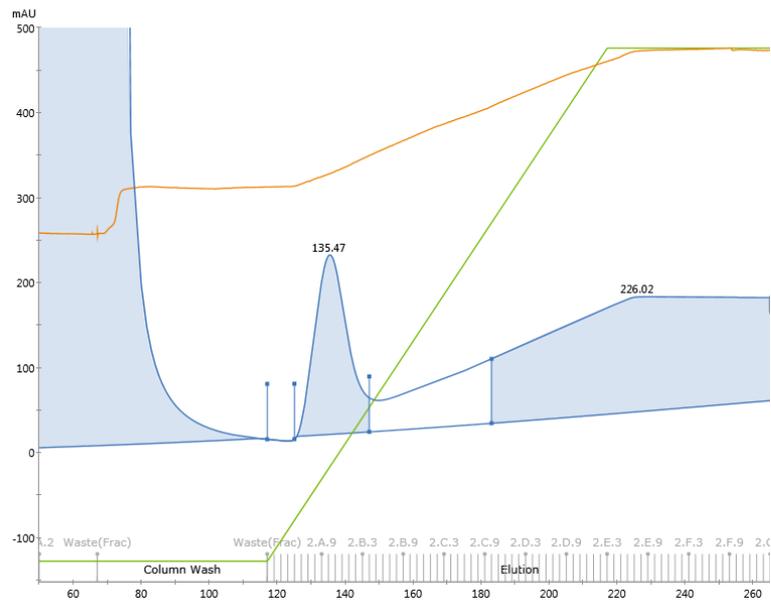


Figure A2.1 Example of UV trace data from affinity chromatography of 6xHis-tagged AtIPK1. Lysed supernatant from expression cultures were loaded onto 5 mL NiNTA column Protein peak fractions highlighted in blue were collected for SDS-PAGE analysis.

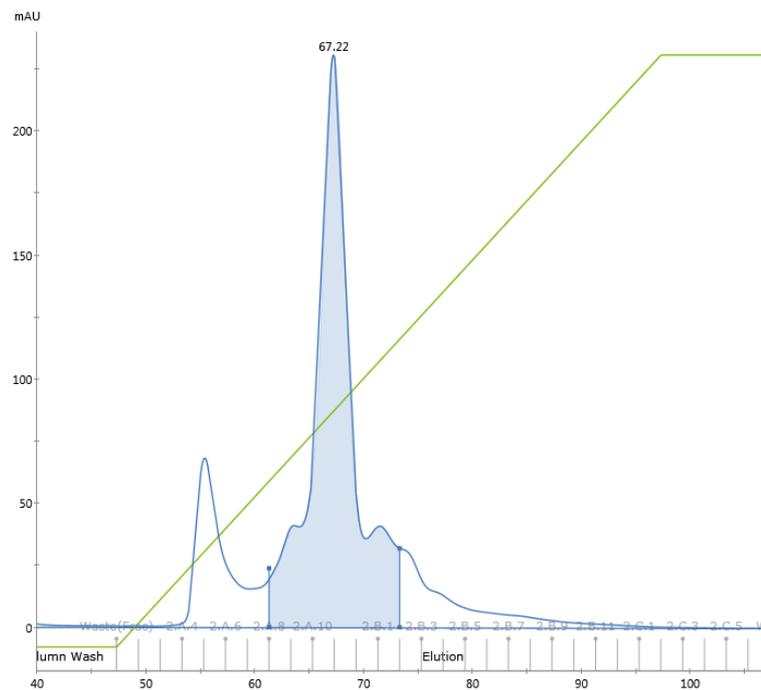


Figure A2.2 Example of UV trace data from heparin chromatography of 6xHis-tagged AtIPK1. Protein from affinity chromatography loaded onto 1 mL heparin column. Protein peak fractions highlighted in blue were collected for SDS-PAGE analysis.

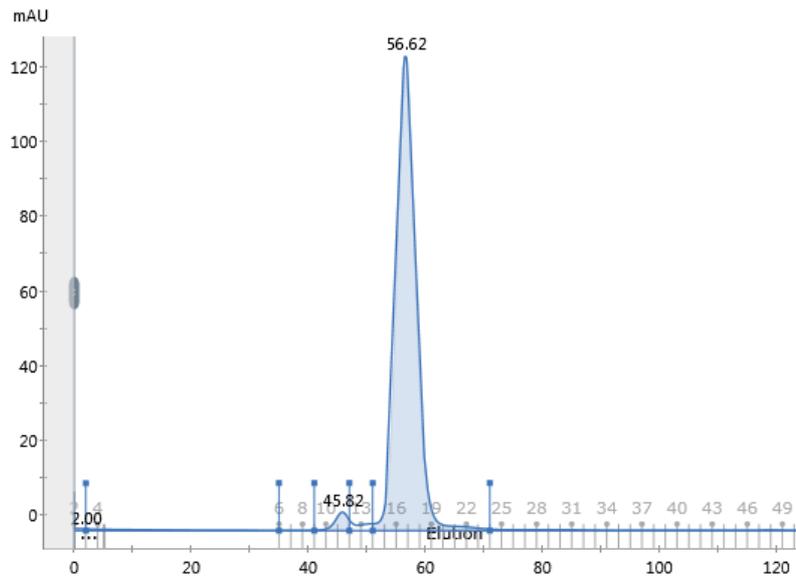
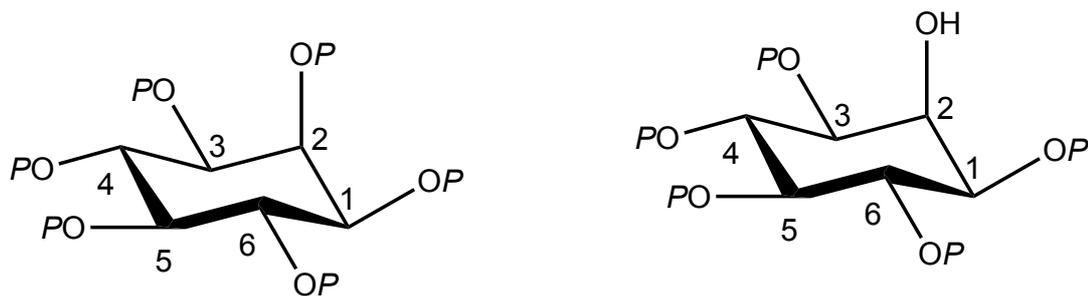


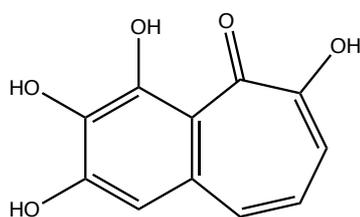
Figure A2.3 Example of UV trace data from size exclusion chromatography of AtIPK1. After 3C protease digest of 6xHis tag and protein concentration, 2 mL of AtIPK1 solution loaded onto Sepharose superdex 75 16/600 column. Protein peak fractions highlighted in blue were collected for SDS-PAGE analysis.

Appendix 3 Ligands for inhibitor assays

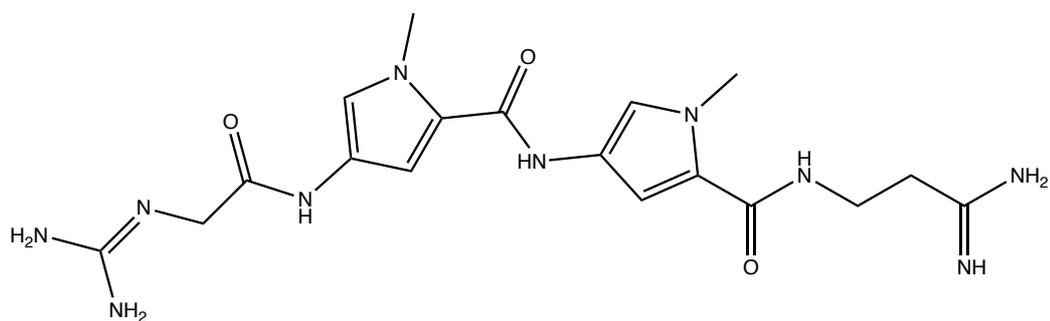


InsP₆

Ins(1,3,4,5,6)P₅



Purpurogallin



Netropsin

Figure A3. Structures of potential inhibitors used in FP displacement assays for AtIPKI-DNA binding. Structures made in ChemDraw.

Appendix 4 Comparison of wildtype AtIPK1-ADP with W129A mutant-ADP structures

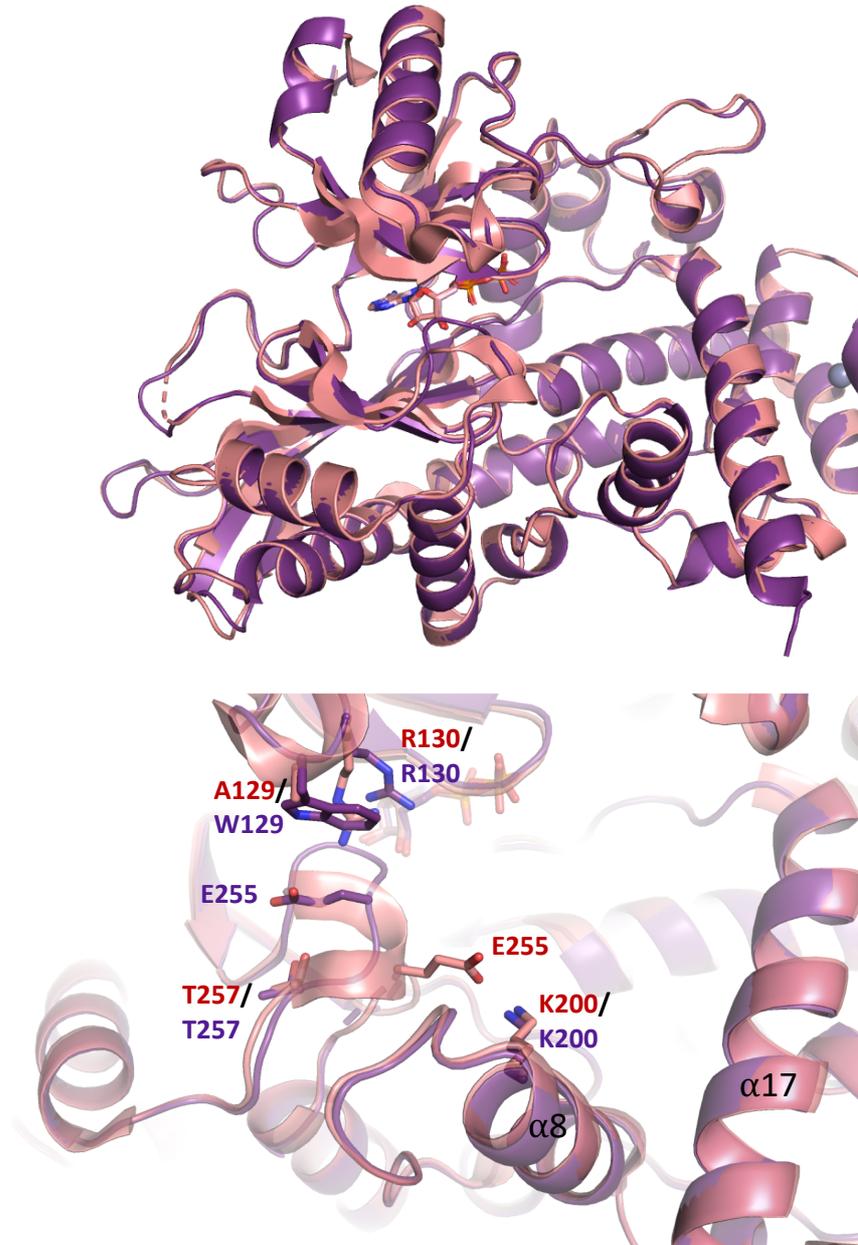


Figure A4 Alignment of wildtype ADP-bound AtIPK1 (purple) with W129A model PDB 4AXE (pink). (a) overall alignment, representing peptide as cartoon, (b) zoomed in image of models highlighting E255-K200 ion pair, characteristic of open and half-open conformations, and residues that stack as sticks.

Appendix 5 Comparison of wildtype apo and ADP bound structures

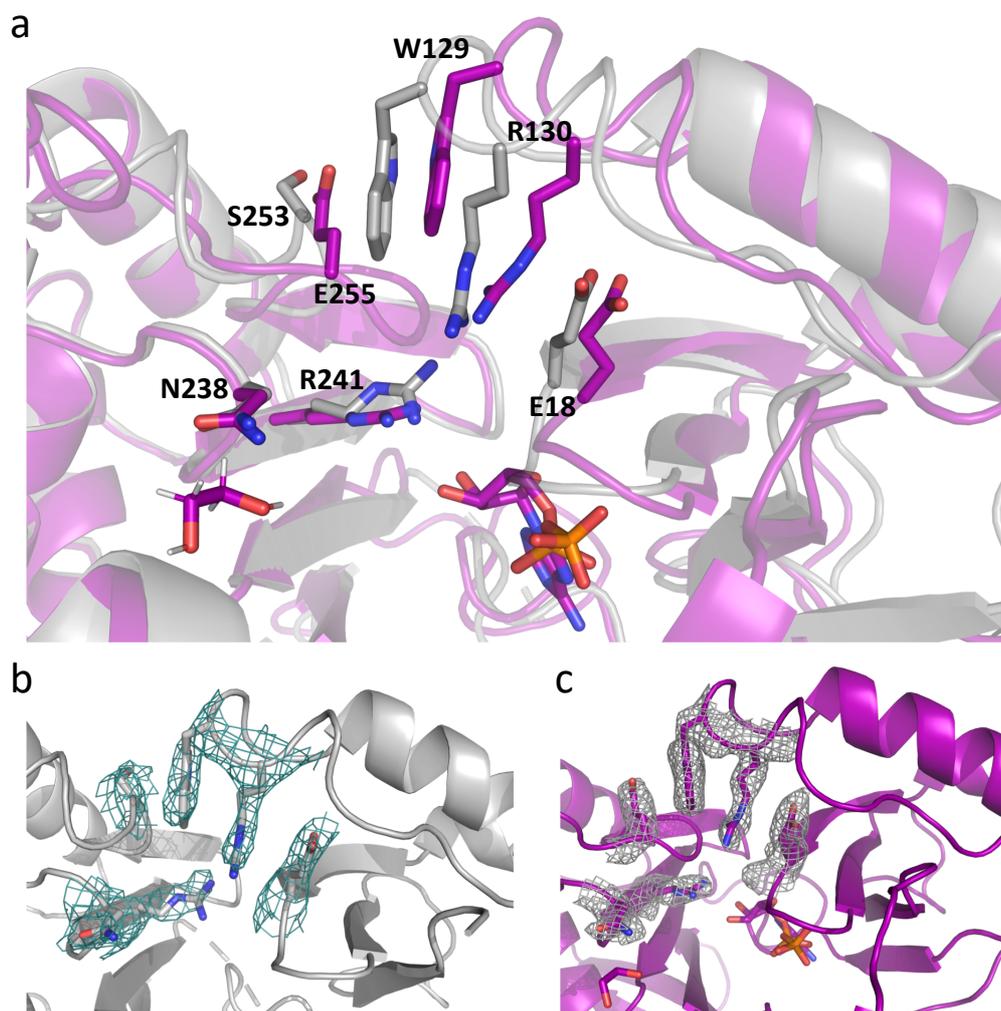


Figure A5 Alignments of wildtype apo (grey) and ADP bound (purple) structures highlighting clasp forming residues as sticks; showing (a) aligned structures; and density maps for (b) apo structure and (c) ADP structure.

Appendix 6 Sequence data of AtIPK1 GRP mutants

A6.1 Example sequence data for AtIPK1 GKP mutant from Mix2seq and ExPASy translate.

DNA_sequence

```
Ttctgcagtgttttcggccttctgcttccgattgtaaagctaattatcttcttatccaatttgtagtatgactc
cattctcttaagtggtttcaggccttagatcaatgaaatgtaccttgaatcaaacgtttgattggctcggttttag
agagacgtaatcaccagagggttctgaatccaagcattccttgattgaaaactgatcataatactacagtcttt
ggcagttgcagctatcaaatactccttcacgatcttcaagctttcatctaaaggtaaagcatgtagagacaattc
cgcctccagtggttaccttctttacatatagggcaaggctggttgataatatcgtaataacaatgaatcgctcc
ttcaatgtctaatttgtctagcttctgaatttcaagaagtctatcaagaactcctgagccatagacagcgtcaga
tactagctgtagaaagcactctgtcctatgaccgtcttctgattgaatgaagcctttgagagcatcctcaaaggc
atacccgatttcggggctggttctcccggctgtttcacctgaacccctaataatgagagaaccattcaagaatac
gcggaaattgttttgaggagtggaatataaagcttttatagcttccaaaactctctctttggatccagagaagag
atcaagaggatcgtactcgcttcttccagatatctcgatatattccaactttaggagttgatgcattttgaaacg
gcttacgcttgttttgagtatgttttctttacctatgaatcttgaggttggaagaaatccgcatttaggctttat
ttcaactaatgcaatcaccaccgctagtaataccttgagagaatagtgaatgatcattcaatatgagagcgga
atcatgactgttatcaacattagctgcatgac
```

3'5' Frame 3

```
cat gca gct aat gtt gat aac agt cat gat tcc gct ctc ata ttg aat gat cat tca cta
H A A N V D N S H D S A L I L N D H S L
ttc tct caa ggt att act agc ggt ggt gat tgc att agt gtt gaa ata aag cct aaa tgc
F S Q G I T S G G D C I S V E I K P K C
gga ttt ctt cca acc tca aga ttc ata ggt aaa gaa aac ata ctc aaa aca agc gta agc
G F L P T S R F I G K E N I L K T S V S
cgt ttc aaa atg cat caa ctc cta aag ttg gaa tat atc gag ata tct gaa gaa agc gag
R F K M H Q L L K L E Y I E I S E E S E
tac gat cct ctt gat ctc ttc tct gga tcc aaa gag aga gtt ttg gaa gct ata aaa gct
Y D P L D L F S G S K E R V L E A I K A
tta tat tcc act cct caa aac aat ttc cgc gta ttc ttg aat ggt tct ctc ata tta ggg
L Y S T P Q N N F R V F L N G S L I L G
ggg tca ggt gaa agc acc ggg aga acc agc ccc gaa atc ggg tat gcc ttt gag gat gct
G S G E S T G R T S P E I G Y A F E D A
ctc aaa ggc ttc att caa tca gaa gac ggt cat agg aca gag tgc ttt cta cag cta gta
L K G F I Q S E D G H R T E C F L Q L V
tct gac gct gtc tat ggc tca gga gtt ctt gat aga ctt ctt gaa att cag aag cta gac
S D A V Y G S G V L D R L L E I Q K L D
aaa tta gac att gaa gga gcg att cat tgt tat tac gat att atc aac cag cct tgc cct
K L D I E G A I H C Y Y D I I N Q P C P
ata tgt aaa gaa ggt aag cca ctg gag gcg gaa ttg tct cta cat gct tta cct tta gat
I C K E G K P L E A E L S L H A L P L D
```

```

gaa agc ttg aag atc gtg aag gag tat ttg ata gct gca act gcc aaa gac tgt agt att
E S L K I V K E Y L I A A T A K D C S I
atg atc agt ttt caa tca agg aat gct tgg gat tca gaa ccc tct ggt gat tac gtc tct
M I S F Q S R N A W D S E P S G D Y V S
cta aaa ccg acc aat caa acg ttt gat tac aag gta cat ttc att gat cta agc ctg aaa
L K P T N Q T F D Y K V H F I D L S L K
cca ctt aag aga atg gag tca tac tac aaa ttg gat aag aag ata att agc ttt tac aat
P L K R M E S Y Y K L D K K I I S F Y N
cgg aag cag aaa gcc gaa aac act gca gaa
R K Q K A E N T A E

```

A6.2 Example sequence data for AtIPK1 SGP mutant from Mix2seq and ExPASy translate.

DNA_sequence

```

ttctgcagtgttttcggcctttctgcttccgattgtaaaagctaattatcttcttatccaatttgtagtatgactc
cattctcttaagtggtttcaggccttagatcaatgaaatgtaccttgaatcaaacgtttgattggctcggttttag
agagacgtaatcaccagagggttctgaatccaagcattccttgattgaaaactgatcataatactacagtcttt
ggcagttgcagctatcaaatactccttcacgatcttcaagctttcatctaaaggtaaagcatgtagagacaattc
cgcctccagtgggcccactttctttacatatagggcaaggctggttgataatatcgtaataacaatgaatcgctcc
ttcaatgtctaattttgtctagcttctgaatttcaagaagtctatcaagaactcctgagccatagacagcgtcaga
tactagctgtagaaagcactctgtcctatgaccgtcttctgattgaatgaagcctttgagagcatcctcaaaggc
ataccgatttcggggctggttctcccggctgttccactgaacccctaatatgagagaaccattcaagaatac
gcggagattgttttgaggagtggaatataaagcttttatagcttccaaaactctctctttggatccagagaagag
atcaagaggatcgtactcgctttcttcagatatctcgatatattccaactttaggagttgatgattttgaaacg
gcttacgcttggtttgagatggtttctttacctatgaatcttgagttggaagaaatccgatttaggctttatt
tcacactaatgcaatcaccaccgctagtaataccttgagagaatagtgaatgatcattcaata

```

3'5' Frame 1

```

tat tga atg atc att cac tat tct ctc aag gta tta cta gcg gtg gtg att gca tta gtg
Y - M I I H Y S L K V L L A V V I A L V
tga aat aaa gcc taa atg cgg att tct tcc aac tca aga ttc ata ggt aaa gaa aac ata
- N K A - M R I S S N S R F I G K E N I
ctc aaa aca agc gta agc cgt ttc aaa atg cat caa ctc cta aag ttg gaa tat atc gag
L K T S V S R F K M H Q L L K L E Y I E
ata tct gaa gaa agc gag tac gat cct ctt gat ctc ttc tct gga tcc aaa gag aga gtt
I S E E S E Y D P L D L F S G S K E R V
ttg gaa gct ata aaa gct tta tat tcc act cct caa aac aat ctc cgc gta ttc ttg aat
L E A I K A L Y S T P Q N N L R V F L N
ggg tct ctc ata tta ggg ggt tca ggt gaa agc acc ggg aga acc agc ccc gaa atc ggg
G S L I L G G S G E S T G R T S P E I G
tat gcc ttt gag gat gct ctc aaa ggc ttc att caa tca gaa gac ggt cat agg aca gag
Y A F E D A L K G F I Q S E D G H R T E

```

```

tgc ttt cta cag cta gta tct gac gct gtc tat ggc tca gga gtt ctt gat aga ctt ctt
C F L Q L V S D A V Y G S G V L D R L L
gaa att cag aag cta gac aaa tta gac att gaa gga gcg att cat tgt tat tac gat att
E I Q K L D K L D I E G A I H C Y Y D I
atc aac cag cct tgc cct ata tgt aaa gaa agt ggg cca ctg gag gcg gaa ttg tct cta
I N Q P C P I C K E S G P L E A E L S L
cat gct tta cct tta gat gaa agc ttg aag atc gtg aag gag tat ttg ata gct gca act
H A L P L D E S L K I V K E Y L I A A T
gcc aaa gac tgt agt att atg atc agt ttt caa tca agg aat gct tgg gat tca gaa ccc
A K D C S I M I S F Q S R N A W D S E P
tct ggt gat tac gtc tct cta aaa ccg acc aat caa acg ttt gat tac aag gta cat ttc
S G D Y V S L K P T N Q T F D Y K V H F
att gat cta agc ctg aaa cca ctt aag aga atg gag tca tac tac aaa ttg gat aag aag
I D L S L K P L K R M E S Y Y K L D K K
ata att agc ttt tac aat cgg aag cag aaa gcc gaa aac act gca gaa
I I S F Y N R K Q K A E N T A E

```

A6.3 Example sequence data for AtIPK1 Del(GRP) mutant from Mix2seq and ExPASy translate.

```
# DNA_sequence
```

```

ttctgcagtgttttcggctttctgcttccgattgtaaaagctaattatcttcttatccaatttgtagtatgactc
cattctcttaagtggtttcaggcttagatcaatgaaatgtaccttgtaatcaaacgtttgattggctcggttttag
agagacgtaatcaccagagggttctgaatccaagcattccttgattgaaaactgatcataatactacagtcttt
ggcagttgcagctatcaaatactccttcacgatcttcaagctttcatctaaaggtaaagcatgtagagacaattc
cgctccagttctttacatataggacaaggctggttgataatatcgtaataacaatgaatcgctccttcaatgct
taatttgctagcttctgaatttcaagaagtctatcaagaactcctgagccatagacagcgtcagatactagctg
tagaaagcactctgtcctatgaccgtcttctgattgaaatgaagcctttgagagcatcctcaaaggcataccgat
ttcggggctggttctcccggctttcacctgaaccccctaatatgagagaaccattcaagaatacgcggaaatt
gttttgaggagtggaatataaaagcttttatagcttcaaaaactctctctttggatccagagaagagatcaagagg
atcgtactcgctttcttccagatatctcgatatattccaactttaggagttgatgattttgaaacggcttacgct
tgttttgagatgttttctttacctatgaatcttgaggttggaagaaatccgcatttaggctttattttcaact
aatgcaatcaccaccgctagtaatacgttgagagaatagtgaatgatcattcaatatgaggagcggaaatcatgac
tggtatcaacatcagctgcattgac

```

```
# 3'5' Frame 2
```

```

tca atg cag ctg atg ttg ata cca gtc atg att ccg ctc ctc ata ttg aat gat cat tca
S M Q L M L I P V M I P L L I L N D H S
cta ttc tct caa cgt att act agc ggt ggt gat tgc att agt gtt gaa ata aag cct aaa
L F S Q R I T S G G D C I S V E I K P K
tgc gga ttt ctt cca acc tca aga ttc ata ggt aaa gaa aac ata ctc aaa aca agc gta
C G F L P T S R F I G K E N I L K T S V

```

agc cgt ttc aaa atg cat caa ctc cta aag ttg gaa tat atc gag ata tct gaa gaa agc
S R F K M H Q L L K L E Y I E I S E E S
gag tac gat cct ctt gat ctc ttc tct gga tcc aaa gag aga gtt ttg gaa gct ata aaa
E Y D P L D L F S G S K E R V L E A I K
gct tta tat tcc act cct caa aac aat ttc cgc gta ttc ttg aat ggt tct ctc ata tta
A L Y S T P Q N N F R V F L N G S L I L
ggg ggt tca ggt gaa agc acc ggg aga acc agc ccc gaa atc ggg tat gcc ttt gag gat
G G S G E S T G R T S P E I G Y A F E D
gct ctc aaa ggc ttc att caa tca gaa gac ggt cat agg aca gag tgc ttt cta cag cta
A L K G F I Q S E D G H R T E C F L Q L
gta tct gac gct gtc tat ggc tca gga gtt ctt gat aga ctt ctt gaa att cag aag cta
V S D A V Y G S G V L D R L L E I Q K L
gac aaa tta gac att gaa gga gcg att cat tgt tat tac gat att atc aac cag cct tgt
D K L D I E G A I H C Y Y D I I N Q P C
cct ata tgt aaa gaa ctg gag gcg gaa ttg tct cta cat gct tta cct tta gat gaa agc
P I C K E L E A E L S L H A L P L D E S
ttg aag atc gtg aag gag tat ttg ata gct gca act gcc aaa gac tgt agt att atg atc
L K I V K E Y L I A A T A K D C S I M I
agt ttt caa tca agg aat gct tgg gat tca gaa ccc tct ggt gat tac gtc tct cta aaa
S F Q S R N A W D S E P S G D Y V S L K
ccg acc aat caa acg ttt gat tac aag gta cat ttc att gat cta agc ctg aaa cca ctt
P T N Q T F D Y K V H F I D L S L K P L
aag aga atg gag tca tac tac aaa ttg gat aag aag ata att agc ttt tac aat cgg aag
K R M E S Y Y K L D K K I I S F Y N R K
cag aaa gcc gaa aac act gca gaa
Q K A E N T A E

Appendix 7 Purification of AtIPK1 GRP mutants

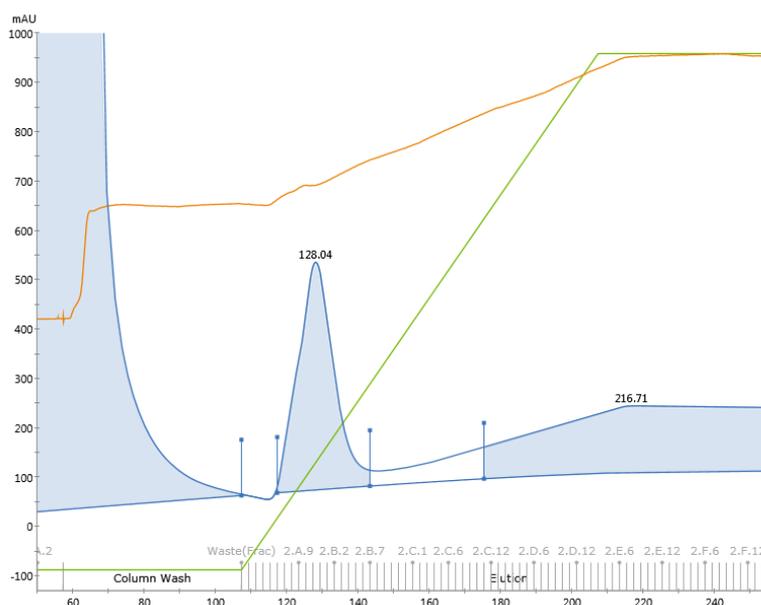


Figure A7.1 Example of UV trace data from affinity chromatography of 6xHis-tagged AtIPK1 mutant. Supernatant from lysed expression cultures were loaded onto 5 mL NiNTA column. Trace data corresponds to purification of GKP mutant. Protein peak fractions highlighted in blue were collected for SDS-PAGE analysis.

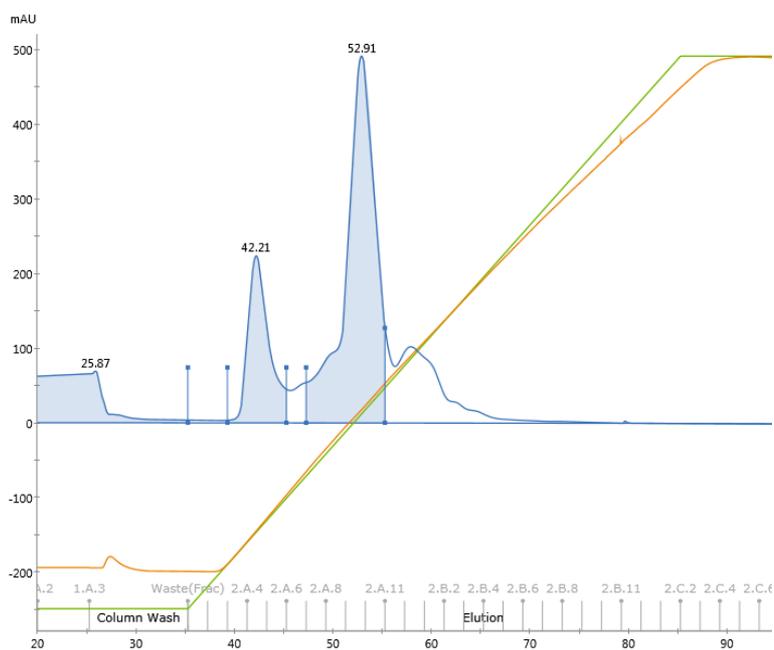


Figure A7.2 Example of UV trace data from heparin chromatography of 6xHis-tagged AtIPK1 mutant. Trace data corresponds to purification of GKP mutant. Protein peak fractions highlighted in blue were collected for SDS-PAGE analysis.

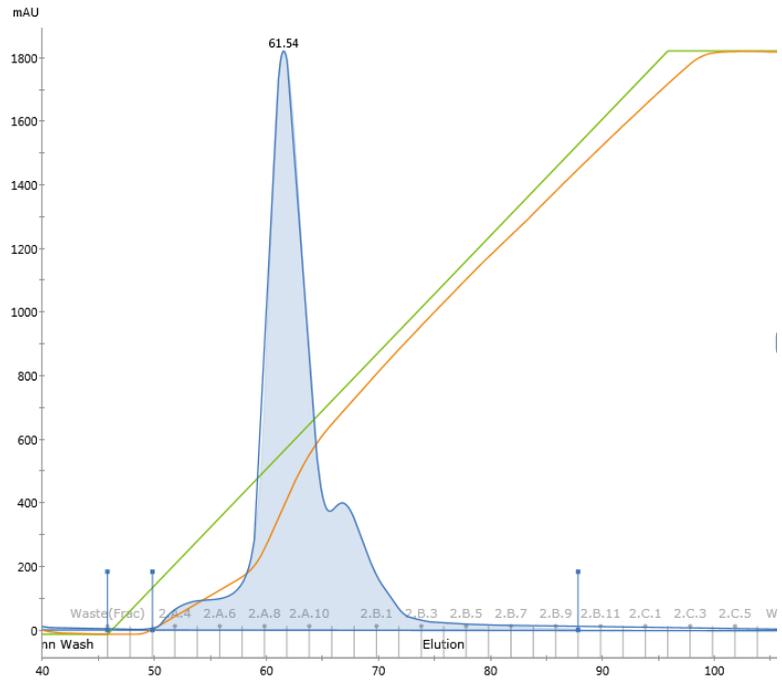


Figure A7.3 Example of UV trace data from size exclusion chromatography of **AtIPK1 mutant**. Trace data corresponds to purification of GKP mutant from Sepharose superdex 75 16/600 column. Protein peak fractions highlighted in blue were collected for SDS-PAGE analysis.

Appendix 8 Electrostatic potential of AtIPK1 surface

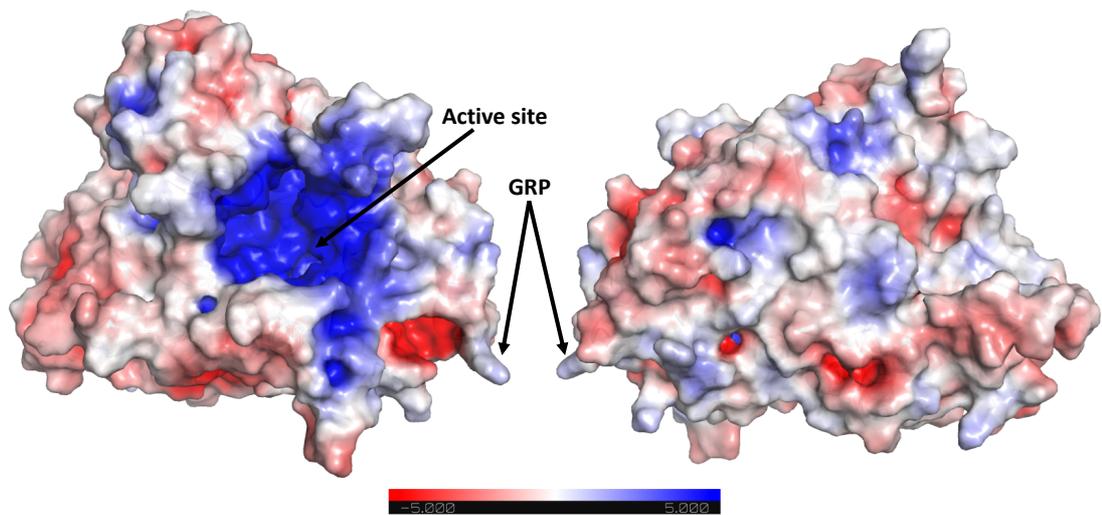
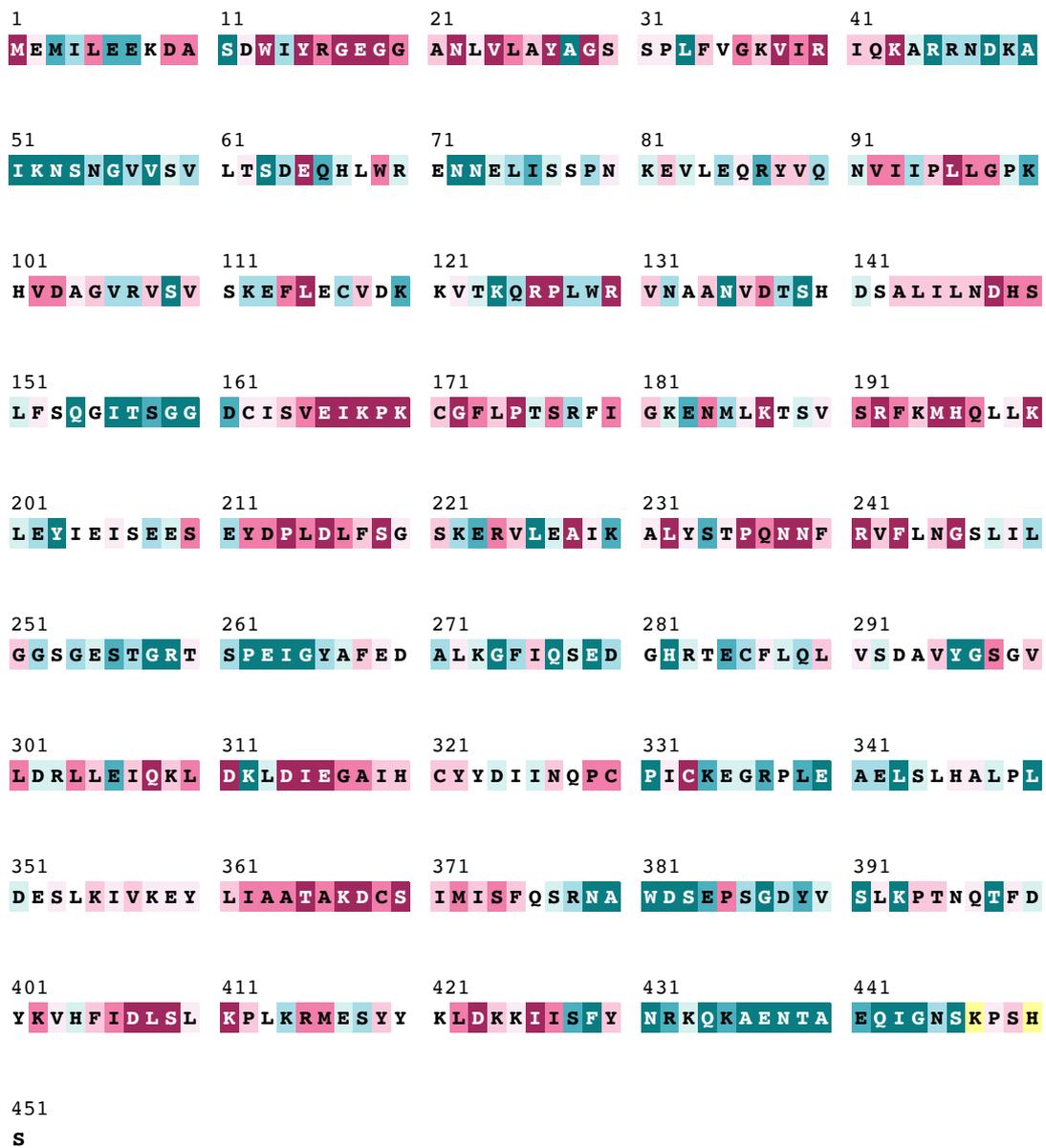


Figure A8 Electrostatic potential across AtIPK1 surface, calculated with APBS plugin PDB2QPR. Two orientations of AtIPK1.

Appendix 9 Conservation of AtIPK1 amino acids



The conservation scale:



X - Insufficient data - the calculation for this site was

Figure A9 ConSurf-DB results showing highly conserved vs. lowly conserved amino acids from AtIPK1 against 470 final homologues.

Appendix 10 Model data from AtIPK1-DNA HDOCK analyses

Table A10.1 Top 10 models of protein-DNA modelling from HDOCK, using Wt AtIPK1 apo structure and [PDB] dodecamer as search models.

Rank	1	2	3	4	5	6	7	8	9	10
Docking Score	-210.16	-210.1	-197.19	-190.26	-184.49	-180.12	-179.95	-177.83	-174.74	-174.18
Ligand rmsd (Å)	94.84	92.61	91.92	102.89	93.55	80.29	81.05	81.26	111.2	94.46
Interface residues	model 1	model 2	model 3	model 4	model 5	model 6	model 7	model 8	model 9	model 10

Row 1: The ranks of the models.

Row 2: The docking energy scores.

Row 3: The ligand RMSDs from the input structures or modelled structures by homology modelling.

Row 4: The interface residues within 5.0 Å from their interacting partner or each other, and the corresponding distances.

Table A10.1.1 Putative DNA binding residues of AtIPK1 from top 10 models

Model 1			Model 2			Model 2 (cont.)			Model 3			Model 4			Model 5			Model 6		
GLU	18	4.029	GLU	18	4.23	TYR	420	3.075	LEU	151	3.465	GLY	181	4.74	GLU	183	4.44	THR	62	3.716
GLY	20	3.311	GLY	19	4.426	LYS	421	2.567	PHE	152	4.329	LYS	182	3.249	LEU	199	3.682	SER	63	2.75
ALA	21	4.477	GLY	20	2.765	LYS	424	3.421	SER	153	2.083	GLU	183	2.397	TYR	323	2.822	ASP	64	2.862
LYS	43	3.441	ALA	21	4.972				ASP	161	2.614	ASN	184	2.475	GLN	328	2.732	HIS	67	2.99
ALA	44	3.197	LYS	43	2.912				CYS	162	4.521	GLU	316	2.396	PRO	329	4.284	ARG	70	2.635
ARG	45	2.082	ALA	44	2.924				ASN	245	3.23	GLN	328	4.261	CYS	330	4.11	TYR	322	3.619
THR	123	4.19	ARG	45	3.103				GLY	274	2.801	PRO	329	4.268	PRO	331	3.462	TYR	323	2.333
LEU	128	3.372	LEU	128	2.634				PHE	275	4.01	PRO	331	4.76	ILE	332	3.786	ILE	326	3.212
TRP	129	2.492	TRP	129	3.202				GLN	277	2.987	ILE	332	4.933	LYS	334	2.384	ASN	327	2.2
ARG	130	2.626	ARG	130	3.54				GLN	376	3.259	LYS	334	2.809	GLU	335	3.5	GLN	328	2.574
VAL	131	3.628	VAL	131	2.663				SER	377	3.525	GLU	335	2.165	GLU	417	2.783	PRO	329	4.827
ASN	132	2.274	ASN	132	2.063				ARG	378	3.763	ARG	337	2.387	SER	418	4.594	PRO	331	3.576
ALA	133	3.248	ALA	133	2.806				ASN	379	2.629	PRO	338	4.115	TYR	420	3.047	LYS	334	3.497
ALA	134	1.548	ALA	134	2.597				ALA	380	2.576	LEU	339	3.579	LYS	421	2.653	PRO	349	4.122
ASN	135	4.436	HIS	196	3.905				TRP	381	2.125	ALA	341	4.704	LYS	424	1.338	LEU	350	2.701
HIS	196	4.824	LEU	199	3.211				ASP	382	2.58	TYR	420	2.401	LYS	425	2.89	ASP	351	3.67
LEU	199	3.577	GLY	254	4.582				ASN	396	4.431	LYS	424	1.591	ILE	426	4.151	LYS	414	3.416
SER	253	3.829	GLU	255	3.122				GLN	397	3.029	LYS	425	4.751	SER	428	2.336	GLU	417	3.328
GLY	254	3.739	SER	256	2.739							SER	428	2.228	PHE	429	3.169			
GLU	255	3.254	ILE	319	4.797							ASN	431	2.661	ARG	432	3.169			
SER	256	2.827	TYR	323	2.182							ARG	432	4.27						
PRO	412	4.761	GLN	328	4.571							GLN	434	3.526						
LYS	414	2.336	PRO	331	3.069															
ARG	415	3.29	ILE	332	3.113															
GLU	417	2.059	PRO	412	4.549															
SER	418	3.337	LYS	414	3.057															
LYS	421	2.789	ARG	415	1.408															
LEU	422	3.8	GLU	417	3.17															
			SER	418	3.838															

Table A10.1.1 (cont.)

Model 7			Model 8			Model 9			Model 10		
GLU	6	2.793	ALA	44	4.753	LYS	182	2.043	HIS	196	3.897
LYS	8	3.12	ARG	45	2.856	GLU	183	2.131	LEU	199	3.017
ASP	9	2.064	LEU	61	2.737	ASN	184	4.848	TYR	322	4.818
SER	11	4.682	THR	62	2.317	MET	185	2.565	TYR	323	1.227
ASP	12	3.4	SER	63	3.174	LEU	186	4.807	GLN	328	2.992
SER	30	3.503	ASP	64	2.87	GLU	316	4.8	PRO	329	3.453
SER	31	4.728	GLU	65	4.843	ILE	332	2.358	PRO	331	3.089
PRO	32	2.902	PRO	79	3.778	CYS	333	4.758	ILE	332	3.503
VAL	35	2.982	ASN	80	3.801	LYS	334	2.942	LYS	334	2.889
GLY	36	2.814	LYS	81	3.73	GLU	335	2.555	GLU	335	4.292
LEU	151	2.687	ASP	119	3.724	GLY	336	4.611	GLU	417	2.705
PHE	152	3.782	THR	123	3.994	ARG	337	2.504	TYR	420	3.336
SER	153	2.835	ASN	132	2.637	PRO	338	3.699	LYS	421	2.962
ASP	161	2.947	ALA	133	3.629	LEU	339	3.672	LEU	422	3.512
CYS	162	4.016	ALA	134	4.405	ALA	341	4.857	LYS	424	3.061
ASN	245	2.953	ASN	135	2.85	TYR	420	2.569	LYS	425	2.317
GLN	376	3.395	VAL	136	4.177	LYS	424	1.986	ILE	426	4.848
ARG	378	3.186	THR	138	3.061	GLN	434	2.611	SER	428	4.497
ASN	379	3.98	SER	139	4.746				PHE	429	2.868
TRP	381	2.627	TYR	322	3.17				ARG	432	2.96
ASP	382	2.876	ILE	325	4.204						
GLY	387	4.623	ILE	326	2.71						
ASP	400	3.474	GLN	328	4.055						
LYS	402	3.746	LEU	410	4.678						
			PRO	412	3.263						
			LEU	413	3.38						
			LYS	414	2.45						
			ARG	415	2.773						
			GLU	417	2.337						

Table A10.2 Top 10 models of protein-DNA modelling from HDOCK, using Wt AtIPKI-ADP structure and [PDB] dodecamer as search models.

Rank	1	2	3	4	5	6	7	8	9	10
Docking Score	-206.56	-203.74	-193.95	-190.32	-187.44	-183.76	-182.61	-181.25	-178.6	-178.25
Ligand rmsd (Å)	71.61	83.17	84.56	71.64	67.29	86.72	67.04	58.3	74.87	85.45
Interface residues	model 1	model 2	model 3	model 4	model 5	model 6	model 7	model 8	model 9	model 10

Row 1: The ranks of the models.

Row 2: The docking energy scores.

Row 3: The ligand RMSDs from the input structures or modelled structures by homology modelling.

Row 4: The interface residues within 5.0 Å from their interacting partner or each other, and the corresponding distances.

Table A10.2.1 Putative DNA binding residues of AtIPK1 from top 10 models

Model 1			Model 2			Model 2 (cont.)			Model 3			Model 4			Model 4 (cont.)			Model 5		
GLY	20	3.984	GLU	18	2.716	LYS	425	3.667	LYS	100	2.44	ALA	44	2.372	GLU	417	2.305	ARG	45	4.012
LYS	43	4.724	GLY	19	2.791	EDO	1D	2.741	HIS	101	3.826	ARG	45	3.774				ASN	47	4.513
ARG	45	3.942	GLY	20	2.826	ADP	1S	3.659	ASN	147	4.919	ASN	47	4.672				ASP	48	4.988
ASN	47	3.057	LYS	43	3.123				PHE	152	3.586	LYS	49	3.808				LYS	49	3.639
TRP	129	2.884	ASN	47	2.844				SER	153	2.851	ALA	50	4.991				ASN	132	3.854
ASN	132	4.266	LEU	128	3.61				GLN	154	2.721	LYS	52	2.9				ALA	133	3.97
ALA	133	4.181	TRP	129	1.603				ASP	293	4.059	ASN	53	3.243				ALA	134	2.952
ALA	134	3.186	ARG	130	2.404				ALA	294	3.855	SER	54	2.532				LYS	170	4.853
HIS	196	3.341	VAL	131	4.806				TYR	296	3.249	ASN	55	2.55				ARG	192	4.771
LYS	200	2.783	ASN	132	3.522				GLY	297	2.677	GLY	56	2.787				HIS	196	2.886
GLU	202	3.107	ALA	133	3.352				SER	298	2.732	VAL	57	3.671				LYS	200	2.929
TYR	203	2.76	ALA	134	3.939				VAL	300	4.493	VAL	58	3.992				GLU	202	3.626
ILE	204	4.897	LYS	168	2.854				ARG	303	2.696	THR	62	3.59				TYR	203	2.987
GLU	205	2.197	HIS	196	2.885				GLN	376	4.215	SER	63	2.767				GLU	205	3.512
ILE	206	3.961	LYS	200	2.463				ARG	378	4.268	ASP	64	2.389				ILE	206	4.9
GLU	255	3.683	TYR	203	2.54				SER	383	2.786	LYS	112	3.47				ASN	238	3.899
SER	256	2.566	ILE	204	2.846				GLU	384	4.278	GLU	116	2.49				SER	256	3.647
THR	257	4.647	GLU	205	2.157				PRO	385	3.922	ALA	134	4.463				SER	418	2.782
SER	418	4.812	ILE	206	4.392				SER	386	2.325	ASN	135	4.636				TYR	419	4.544
LEU	422	3.546	SER	207	2.385				GLY	387	3.001	ASP	137	2.653				LEU	422	4.261
LYS	425	2.323	GLU	208	4.854				ASP	388	1.666	SER	139	3.903				LYS	424	4.433
SER	428	4.139	ASN	238	2.417				TYR	389	2.613	HIS	140	4.638				LYS	425	3.334
PHE	429	4.354	ARG	241	3.706				VAL	390	4.321	TYR	322	3.985				SER	428	2.219
ARG	432	2.301	GLY	254	3.28				SER	391	3.211	ILE	326	4.577				PHE	429	3.315
LYS	433	3.36	GLU	255	3.746				TYR	401	2.206	ASN	327	4.343				ARG	432	2.504
			THR	257	4.977				LYS	402	3.1	GLN	328	3.401				EDO	1D	2.485
			GLY	258	3.234							PRO	329	3.9						
			ARG	259	4.003							LYS	334	4.683						
			ASP	407	4.461							LYS	414	2.654						

Table A10.2.1 cont.

Model 6			Model 7			Model 7 (cont.)			Model 8			Model 9			Model 10		
TYR	15	3.617	GLY	19	4.793	SER	418	2.913	TYR	203	4.446	GLU	18	4.103	ALA	44	3.867
ARG	16	4.546	GLY	20	3.082	TYR	419	4.141	TYR	322	4.379	GLY	19	4.051	ARG	45	2.64
LYS	124	4.778	ALA	21	4.969	LYS	421	4.808	TYR	323	1.941	GLY	20	3.133	ASN	47	3.395
GLN	125	3.587	LYS	43	2.838	LEU	422	4.092	ILE	326	4.184	LYS	43	3.974	GLU	116	4.418
ARG	126	3.207	ALA	44	4.947	LYS	424	3.308	GLN	328	3.136	LYS	49	2.938	ASP	119	1.614
PRO	127	2.972	ARG	45	3.164	LYS	425	2.552	PRO	329	2.736	ARG	126	4.702	LYS	120	3.097
LEU	128	2.592	ASN	47	2.919	SER	428	2.644	CYS	330	4.812	LEU	128	3.443	THR	123	2.6
TRP	129	2.901	ASP	48	4.068	PHE	429	4.634	PRO	331	3.267	TRP	129	2.556	TRP	129	2.486
ASN	132	2.927	LYS	49	2.687	ARG	432	2.573	ILE	332	4.112	ARG	130	3.102	VAL	131	3.607
ALA	133	3.071	ALA	50	3.893				LYS	334	2.555	VAL	131	3.102	ASN	132	2.463
ALA	134	4.773	TRP	129	4.8				GLU	335	2.573	ASN	132	1.422	ALA	133	3.891
TYR	203	3.221	ASN	132	4.926				GLU	417	2.421	ALA	133	1.694	ALA	134	3.629
ILE	204	2.68	ALA	133	3.917				TYR	420	3.111	ALA	134	3.21	ASN	135	2.578
GLU	205	0.548	ALA	134	2.512				LYS	421	3.584	LYS	168	4.49	VAL	136	3.217
ILE	206	2.768	ARG	192	4.638				LYS	424	2.031	LYS	200	3.771	THR	138	4.476
SER	207	2.065	HIS	196	2.483				LYS	425	2.937	GLU	202	4.315	HIS	196	4.36
PHE	243	4.824	LYS	200	3.426				SER	428	3.447	TYR	203	2.523	TYR	203	2.429
SER	247	4.733	GLU	202	4.722				ARG	432	2.793	ILE	204	2.408	ILE	204	2.826
LEU	248	2.707	TYR	203	2.994							GLU	205	1.916	GLU	205	2.538
GLY	251	2.944	GLU	205	1.565							ILE	206	4.165	GLY	254	3.779
GLY	252	4.262	ILE	206	2.547							ASN	238	3.129	GLU	255	4.165
SER	253	3.548	GLN	237	4.993							GLY	254	3.969	SER	256	4.947
GLU	255	2.564	ASN	238	4.64							GLU	255	2.962	LEU	422	2.999
SER	256	2.036	GLU	255	4.358							SER	256	2.988	LYS	425	4.068
THR	257	3.688	SER	256	2.815							LYS	425	3.939			
GLY	258	3.484	LYS	334	4.887							EDO	1D	2.093			
GLU	263	3.289	GLU	335	2.691												
ILE	264	4.324	GLY	336	4.855												
			ARG	415	2.937												

Table A10.3 Top 10 models of protein-DNA modelling from HDOCK, using Wt AtIPKI-ADP-InsP₆ structure (PDB 2XAM) and dodecamer (PDB 3UXW) as search models.

Docking Score	-230.34	-214.68	-211.32	-203.76	-202.01	-193.51	-191.36	-189.92	-186.8	-181.88
Ligand rmsd (Å)	90.84	88.2	81.46	79.1	86.88	79.5	88.21	89.88	89.29	79.36
Interface residues	model 1	model 2	model 3	model 4	model 5	model 6	model 7	model 8	model 9	model 10

Row 1: The ranks of the models.

Row 2: The docking energy scores.

Row 3: The ligand RMSDs from the input structures or modelled structures by homology modelling.

Row 4: The interface residues within 5.0 Å from their interacting partner or each other, and the corresponding distances.

Table A10.3.1 Putative DNA binding residues of AtIPKI from top 10 models

Model 1			Model 2			Model 3			Model 4			Model 5			Model 6		
ARG	45	2.286	ARG	45	3.26	GLY	20	4.728	ARG	46	3.301	ARG	46	2.273	ARG	45	4.947
ARG	46	2.562	ARG	46	1.768	ARG	45	4.511	ASN	47	3.387	ASN	47	4.252	ARG	46	3.455
ASN	47	2.889	ASN	47	2.886	ARG	46	3.106	TRP	129	2.595	LYS	49	2.619	ASN	47	2.434
ASP	48	2.828	ASP	48	2.781	ASN	47	2.751	ASN	132	3.34	ASP	119	2.801	ASP	48	4.935
LYS	49	2.795	LYS	120	4.057	ASP	48	4.766	ALA	133	4.121	LYS	120	4.257	TRP	129	3.055
ASP	119	3.748	THR	123	2.59	TRP	129	2.73	HIS	196	2.825	THR	123	4.728	ARG	130	4.28
THR	123	2.622	LYS	124	3.521	ARG	130	3.344	LYS	200	3.626	TRP	129	3.172	ASN	132	2.677
LEU	128	4.887	LEU	128	4.698	ASN	132	3.313	LEU	201	4.335	VAL	131	4.504	ALA	133	3.023
TRP	129	3.89	TRP	129	3.725	ALA	133	3.445	GLU	202	2.688	ASN	132	2.711	GLU	202	3.558
VAL	131	3.538	ARG	130	4.031	LEU	199	3.238	TYR	203	3.106	ASN	135	3.069	TYR	203	2.922
ASN	132	3.22	VAL	131	3.042	GLU	202	3.454	ILE	204	3.593	LEU	199	3.625	GLU	205	4.534
ALA	133	4.507	ASN	132	2.592	TYR	203	1.883	GLU	205	2.469	GLU	202	2.457	GLU	335	3.555
ASN	135	3.487	ALA	133	4.04	SER	418	4.016	ASN	238	4.532	TYR	203	3.084	LYS	424	2.821
TYR	203	3.569	ASN	135	2.638	LYS	424	3.197	SER	418	4.913	GLU	205	2.626	LYS	425	3.727
GLU	205	3.663	TYR	203	2.92	LYS	425	3.042	LYS	421	4.528	LYS	421	4.418	SER	428	3.382
PRO	412	4.045	ILE	204	3.556	ILE	427	4.99	LEU	422	3.943	LYS	424	4.187	PHE	429	3.48
LYS	414	3.703	GLU	205	2.387	SER	428	1.558	LYS	425	3.433	LYS	425	2.878	ARG	432	2.32
ARG	415	2.1	LYS	414	3.552	PHE	429	3.161	PHE	429	2.921	ILE	426	4.924			
SER	418	2.951	ARG	415	3.33	ARG	432	3.223	ARG	432	3.053	SER	428	2.41			
LYS	421	3.226	SER	418	4.222				LYS	433	3.261	PHE	429	2.667			
LYS	425	2.974	LYS	425	2.006							ARG	432	2.856			

Table A10.3.1 (cont.)

Model 7			Model 8			Model 9			Model 10		
ARG	46	2.953	GLY	20	3.758	ARG	45	2.935	ARG	46	3.266
ASN	47	3.991	ARG	45	2.897	ARG	46	2.582	ASN	47	4.346
ASP	119	3.413	ARG	46	2.458	ASN	47	2.57	LYS	49	4.736
LYS	120	4.15	ASN	47	2.737	ASP	48	1.873	TRP	129	2.817
THR	123	2.674	ASP	48	3.155	LYS	49	2.664	ASN	132	2.851
LYS	124	3.475	THR	123	1.872	TRP	129	4.829	ALA	133	4.642
TRP	129	3.324	LYS	124	4.179	ARG	130	4.638	LEU	199	4.349
VAL	131	3.206	LEU	128	3.69	ASN	132	3.45	GLU	202	3.469
ASN	132	2.96	TRP	129	3.179	ALA	133	4.866	TYR	203	3.207
ASN	135	2.568	ARG	130	4.529	ASN	135	3.677	GLU	205	2.745
GLU	202	3.713	VAL	131	3.013	TYR	203	3.915	GLU	335	4.228
TYR	203	2.358	ASN	132	2.46	GLU	205	2.519	LYS	421	2.542
ILE	204	3.197	ASN	135	4.724	PRO	412	3.504	LEU	422	2.985
GLU	205	2.24	GLU	202	4.922	LYS	414	2.804	LYS	424	2.725
LYS	421	4.451	TYR	203	3.309	ARG	415	1.487	LYS	425	2.842
LYS	425	2.76	ILE	204	4.554	SER	418	2.68	SER	428	2.433
PHE	429	4.77	GLU	205	2.237	LYS	421	2.83	PHE	429	2.39
			ARG	415	2.841	LYS	425	3.488	ARG	432	0.872
			SER	418	4.927						
			LYS	425	3.299						

Table 10.4 Top 10 models of protein-DNA modelling from HDOCK, using Wt AtIPK1-ADP-InsP₆ structure (PDB 6FJK) and dodecamer (PDB 3UXW) as search models.

Rank	1	2	3	4	5	6	7	8	9	10
Docking Score	-218	-203.57	-201.83	-198.19	-191.93	-191.68	-191.17	-189.04	-188.52	-187.67
Ligand rmsd (Å)	54.07	56.14	12.29	58.06	26.22	57.46	55.18	49.87	55.42	55.94
Interface residues	model 1	model 2	model 3	model 4	model 5	model 6	model 7	model 8	model 9	model 10

Row 1: The ranks of the models.

Row 2: The docking energy scores.

Row 3: The ligand RMSDs from the input structures or modelled structures by homology modelling.

Row 4: The interface residues within 5.0 Å from their interacting partner or each other, and the corresponding distances.

Table A10.4.1 Putative DNA binding residues of AtIPK1 from top 10 models

Model 1			Model 2			Model 3			Model 3 (cont.)			Model 4			Model 5			Model 6		
ARG	45	2.688	ARG	45	3.207	PRO	32	3.655	LYS	393	4.502	ARG	45	3.178	PRO	0	3.207	ILE	319	4.781
ARG	46	2.894	ARG	46	2.643	LEU	33	4.343	ASN	396	3.374	ARG	46	2.891	LYS	37	3.368	TYR	323	2.44
ASN	47	2.76	ASN	47	2.841	LYS	37	3.654	THR	398	4.621	ASN	47	2.748	GLN	90	3.046	ASP	324	4.652
TRP	129	2.955	TRP	129	2.789	GLN	90	4.957	ASP	400	2.385	TRP	129	2.849	ILE	94	3.72	ILE	326	4.788
ARG	130	4.575	ARG	130	4.601	ILE	94	4.181	TYR	401	3.675	ASN	132	2.993	PRO	99	4.591	ASN	327	2.202
ASN	132	3.118	ASN	132	2.941	GLY	98	4.669	LYS	402	2.309	ALA	133	2.661	LYS	100	2.794	GLN	328	2.825
ALA	133	3.101	ALA	133	3.123	PRO	99	1.45	HIS	404	2.516	ALA	134	4.641	VAL	102	4.516	PRO	329	2.803
GLU	202	2.143	HIS	196	4.466	LYS	100	3.161				HIS	196	4.991	ASP	103	2.237	PRO	331	2.738
TYR	203	2.337	LEU	199	4.96	HIS	101	3.453				TYR	203	3.07	ALA	104	4.646	ILE	332	3.691
GLU	205	3.012	LYS	200	4.485	VAL	102	2.983				ILE	204	3.708	VAL	106	4.622	LYS	334	2.767
LYS	334	4.597	GLU	202	1.551	ASP	103	2.475				GLU	205	2.672	ASN	147	3.257	HIS	346	4.54
SER	418	3.148	TYR	203	1.914	VAL	106	3.62				PRO	331	4.774	LEU	151	3.406	LYS	414	3.258
LYS	424	2.364	GLU	205	2.787	ASN	147	3.075				LYS	414	3.136	PHE	152	2.543	GLU	417	2.915
LYS	425	3.945	SER	418	2.945	ASP	148	4.294				GLU	417	3.888	SER	153	2.985	TYR	420	2.87
SER	428	3.041	LYS	424	3.923	HIS	149	4.202				SER	418	3.15	GLN	154	2.877	LYS	421	3.231
PHE	429	3.43	LYS	425	3.572	LEU	151	3.815				LYS	421	4.34	SER	158	2.669	LYS	424	3.369
ARG	432	2.412	SER	428	2.268	PHE	152	1.537				LYS	424	2.849	GLY	159	2.914	LYS	425	3.068
LYS	433	4.967	PHE	429	3.304	SER	153	3.83				LYS	425	2.032	GLN	376	2.814	SER	428	2.345
			ARG	432	2.943	GLN	154	3.861							SER	377	4.866	PHE	429	4.837
						ASP	293	3.539							ASP	388	4.657	ARG	432	2.403
						TYR	296	3.442							TYR	389	3.358			
						GLY	297	1.667							ASP	400	2.322			
						SER	298	3.701							LYS	402	1.993			
						GLN	376	4.18							VAL	403	3.895			
						ASP	388	2.234							HIS	404	2.034			
						TYR	389	2.357												
						VAL	390	3.317												
						SER	391	2.578												
						LEU	392	4.631												

Table A10.4.1 cont.

Model 7			Model 8			Model 9			Model 10		
GLY	20	4.436	PRO	175	4.074	ARG	45	2.028	ARG	45	2.059
ARG	45	2.552	THR	176	2.641	ARG	46	2.125	ARG	46	2.577
ARG	46	2.249	SER	177	4.219	ASN	47	2.318	ASN	47	3.589
ASN	47	3.842	ARG	178	2.759	ASP	48	3.306	ASP	119	3.345
ASP	119	4.98	PHE	179	3.67	ASP	119	4.648	LYS	120	4.34
THR	123	2.864	ILE	180	4.27	THR	123	3.494	THR	123	2.693
LYS	124	2.291	GLY	181	3.526	LYS	124	4.486	LYS	124	3.043
LEU	128	2.621	LYS	182	3.313	LEU	128	2.922	ARG	126	4.903
TRP	129	2.424	GLU	183	2.716	TRP	129	4.55	LEU	128	4.212
ARG	130	2.596	ASN	184	4.449	VAL	131	3.137	TRP	129	2.842
VAL	131	3.079	MET	185	3.233	ASN	132	2.388	VAL	131	3.274
ASN	132	2.391	THR	188	4.772	ALA	133	4.782	ASN	132	3.368
ALA	133	4.572	GLU	316	3.239	TYR	203	3.034	ASN	135	4.509
TYR	203	2.404	HIS	320	2.485	GLU	205	2.443	GLU	202	3.736
ILE	204	2.484	CYS	330	4.992	LYS	414	4.459	TYR	203	2.338
GLU	205	3.127	ILE	332	3.329	ARG	415	2.865	ILE	204	3.316
LYS	425	3.96	CYS	333	1.75	SER	418	2.9	GLU	205	3.242
			LYS	334	3.354	LYS	421	3.997	SER	418	4.343
			GLU	342	2.509	LYS	425	3.063	LYS	425	3.958
			LEU	343	4.273						
			SER	344	4.937						
			HIS	346	4.791						
			TYR	420	1.853						
			LYS	424	4.508						

Appendix 11 Similar models of AtIPK1 homodimer generated from two *in silico* modelling methods

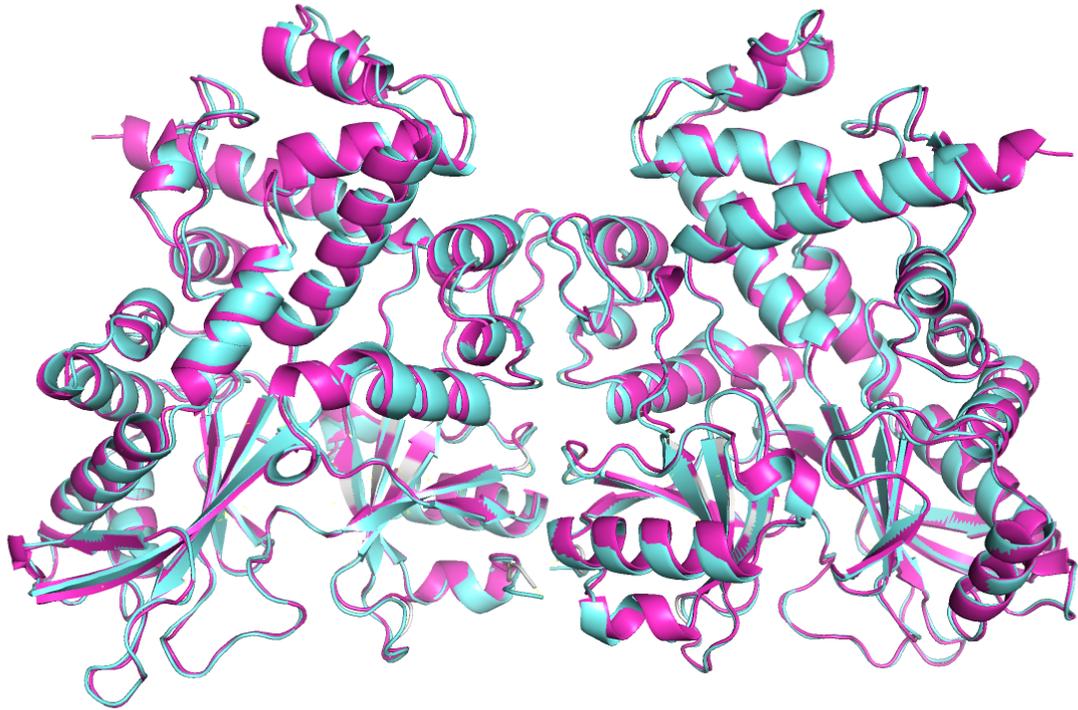


Figure All Alignment of AtIPK1 dimer models from GalaxyHomomer (pink) and HDOCK (aquamarine). RMSD = 0.640 (858 to 858 common atoms).

Appendix 12 Expression data for full length Arabidopsis SPX domain proteins in *E. coli*

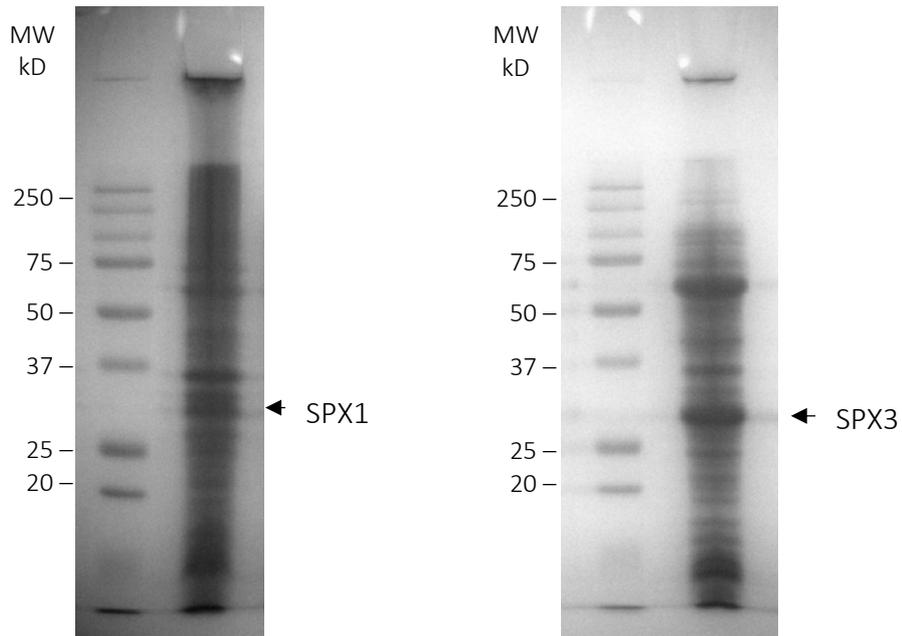


Figure A12.1 SDS-PAGE analysis of 1 L expression pellets for AtSPX1 (~30 kDa) and AtSPX3 (~ 28 kDa) from ArcticExpress RIL.

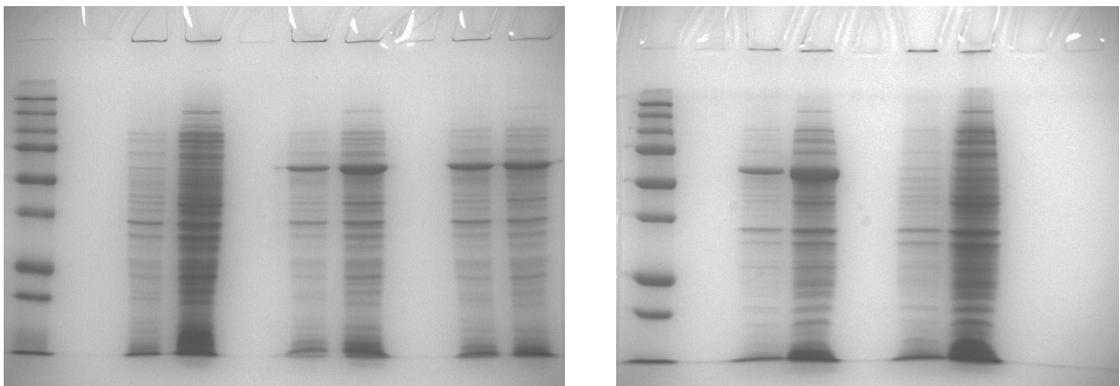


Figure A12.2 No expression of AtPHT5;1. Example of expression trials in ArcticExpress RP induced with 1 mM IPTG 5x uninduced/induced. Cpn60 expression. No AtPHT5.1 (~ 78 kDa).

Appendix 13 Purification of AtSPX3

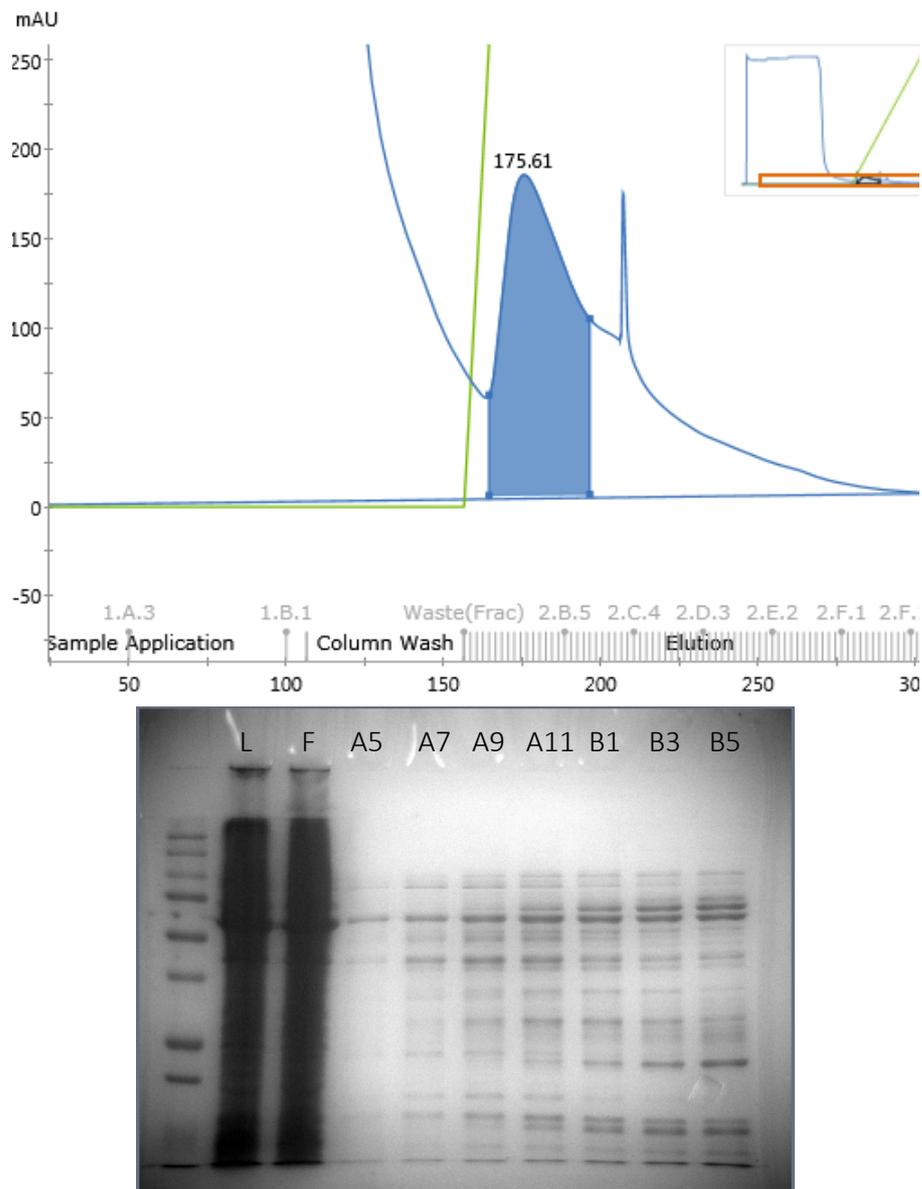


Figure A13 Loss of AtSPX3 during first purification step. (a) UV trace data from affinity chromatography of 6xHis-tagged AtSPX3. Supernatant from lysed expression cultures were loaded onto 5 mL NiNTA column and run at 0.5 mL/min. Protein peak fractions highlighted in blue were collected for SDS-PAGE analysis. (b) SDS-PAGE analysis of AtSPX3 throughout affinity chromatography. Showing L loaded onto column, F column flowthrough, and fractions A5-B5.

Appendix 14 Sequence data verification of cloning Arabidopsis SPX domain proteins into *pOPIN*F

A14.1 Example sequence data for *AtSPX1* in *pOPIN*F from Mix2seq and ExPASy translate.

DNA_sequence

```
cggagttatccgggacctttaattcaaccaacacaatatattatagttaaataaga
attattatcaaatcatttgtatattaattaaatactatactgtaaattacatttta
tttacaatcaaaggagatataccatggcacaccatcaccaccatcacagcagcggtc
tggagttctgtttcagggcccgatgaagtttggttaagagtctcagcaatcagatcg
agcaaactcttcctgaatggcaagacaagttcttgtcttacaaagaactcaaaaac
gactcaaactcatcggttccaaaaccgccgatcgtcccgttaaacgactccgtttag
atgagttttcgtcggaatatcgaaagaagagatcaatttcatccaattgtagaag
acgagttggagaaattcaacaatttcttcgttgagaaggaagaatataatcatca
gactaaaggaatttagagatagaattgcaagactaaggattcaatggagaagatga
taaaaatcaggaaggagattggtgatttccatggagaaatggttcttcttgagaatt
acagtgtcttaattacactggattgggttaagatactgaagaagatgacaaaagaa
ctggtgatctcatgctgtttacctttcatccagaaagttcttcagcaacctttttaca
ctactgacttattgttcaagcttgtcaaggaatctgaggcaatgcttgatcagatct
tcctgctaacgaaactgagtctgagattatccaagcagagttatcagagcataagt
tcatggagagtcttcataatgaagagcaaatcgctgccttgccgggttttgaaggaga
tcaggagtggagatttctactgttagtgtgttttcattgccctctacagttaaatg
gcttagatgagacatggaagaagattccattgttggagcaagaagccaaataaagct
ttctagaccagtttgtgattaacctcaggtgcaggctgcctatca
```

3'5' Frame 2

```
cggag tta tcc ggg acc ttt aat tca acc caa cac aat ata tta tag tta aat aag aat tat
E L S G T F N S T Q H N I L - L N K N Y
tat caa atc att tgt ata tta att aaa ata cta tac tgt aaa tta cat ttt att tac aat
Y Q I I C I L I K I L Y C K L H F I Y N
caa agg aga tat acc atg gca cac cat cac cac cat cac agc agc ggt ctg gaa gtt ctg
Q R R Y T M A H H H H H H S S G L E V L
ttt cag ggc ccg atg aag ttt ggt aag agt ctc agc aat cag atc gag caa act ctt cct
F Q G P M K F G K S L S N Q I E Q T L P
gaa tgg caa gac aag ttc ttg tct tac aaa gaa ctc aaa aaa cga ctc aaa ctc atc ggt
E W Q D K F L S Y K E L K K R L K L I G
tcc aaa acc gcc gat cgt ccc gtt aaa cga ctc cgt tta gat gag ttt tcc gtc gga ata
S K T A D R P V K R L R L D E F S V G I
tcg aaa gaa gag atc aat ttc atc caa ttg tta gaa gac gag ttg gag aaa ttc aac aat
S K E E I N F I Q L L E D E L E K F N N
ttc ttc gtt gag aag gaa gaa gaa tat atc atc aga cta aag gaa ttt aga gat aga att
F F V E K E E E Y I I R L K E F R D R I
gcg aaa gct aag gat tca atg gag aag atg ata aaa atc agg aag gag att gtt gat ttc
A K A K D S M E K M I K I R K E I V D F
cat gga gaa atg gtt ctt ctt gag aat tac agt gct ctt aat tac act gga ttg gtt aag
H G E M V L L E N Y S A L N Y T G L V K
ata ctg aag aag tat gac aaa aga act ggt gat ctc atg cgt tta cct ttc atc cag aaa
I L K K Y D K R T G D L M R L P F I Q K
gtt ctt cag caa cct ttt tac act act gac tta ttg ttc aag ctt gtc aag gaa tct gag
V L Q Q P F Y T T D L L F K L V K E S E
gca atg ctt gat cag atc ttc cct gct aac gaa act gag tct gag att atc caa gca gag
A M L D Q I F P A N E T E S E I I Q A E
tta tca gag cat aag ttc atg gag agt ctt cat atg aag agc aca atc gct gcc ttg cgg
L S E H K F M E S L H M K S T I A A L R
ggt ttg aag gag atc agg agt gga agt tct act gtt agt gtg ttt tca ttg ccg cct cta
V L K E I R S G S S T V S V F S L P P L
cag tta aat ggc tta gat gag aca tgg aag aag att cca ttg ttg gag caa gaa gcc aaa
Q L N G L D E T W K K I P L L E Q E A K
taa agc ttt cta gac cag ttt gtg att aac ctc agg tgc agg ctg cct atc
- S F L D Q F V I N L R C R L P I
```

A14.2 Example sequence data for *AtSPX3* in *pOPIN* from Mix2seq and ExPASy translate.

DNA_sequence

```
aattatccgggacctttaattcaaccaacacaatatattatagttaaataagaatt
attatcaaatcatttgtatattaattaaaatactatactgtaaattacattttattt
acaatcaaaggagatataccatggcacacccatcaccacccatcacagcagcgggtctgg
aagttctgttttcagggcccgatgaagtttggaagaggattaaagaacagatacaag
agtcgttgccggagtgccgagacaagtttcttcgttacaaggaactcaagaatctga
tctcttctccggcgccggtggaatctatcttctcgtcggtttggtgaacgcagagatcg
acaagtttaatgctttcttcgctgaacaagaagaagatttcatcatccaccacaagg
agttgcaatatcggattcagagattggttagagaaatgtggacacaatgatgaaatgt
ctagagagaatattagtgagatcagaaaagatattgtcaatttccatggcgaatgg
ttctgctagtaaacctacagtaacatcaattacactggattagcaaagattctaaaga
agtacgacaagcgaacaagaggaggattaagatcaccatttattcaaaaagttcttc
atcaaccgtttttcaagactgatcttgtttcaagactagtaagagagtgggagacga
cgatggacgcggtggatccagtgaaaggtggcggaggcggagggatacagagagatgtg
cggcggtgacttcggcagcggcgggagaagggatatttaggaatacggttgcggcat
tattgactatgaaagagatgagaagaggaagttcgacttacagtgcatctcacttc
cgccgctaaatatctccgattccgataatgttctccgatctcttcatctatcttctc
cgattcctattccataaagctttctagaccagtttgtgattaacctcaggtgcaggc
tgcttatcagaaggtggtggctggt
```

3'5' Frame 3

```
aatta tcc ggg acc ttt aat tca acc caa cac aat ata tta tag tta aat aag aat tat tat
  L S G T F N S T Q H N I L - L N K N Y Y
caa atc att tgt ata tta att aaa ata cta tac tgt tgt aaa tta cat ttt att tac aat caa
  Q I I C I L I K I L Y C K L H F I Y N Q
agg aga tat acc atg gca cac cat cac cac cat cac agc agc ggt ctg gaa gtt ctg ttt
  R R Y T M A H H H H H S S G L E V L F
cag ggc ccg atg aag ttt gga aag agg att aaa gaa cag ata caa gag tcg ttg ccg gag
  Q G P M K F G K R I K E Q I Q E S L P E
tgg cga gac aag ttt ctt cgt tac aag gaa ctc aag aat ctg atc tct tct ccg gcg ccg
  W R D K F L R Y K E L K N L I S S P A P
gtg gaa tct att ttc gtc ggt ttg ttg aac gca gag atc gac aag ttt aat gct ttc ttc
  V E S I F V G L L N A E I D K F N A F F
gtc gaa caa gaa gaa gat ttc atc atc cac cac aag gag ttg caa tat cgg att cag aga
  V E Q E E D F I I H H K E L Q Y R I Q R
ttg gta gag aaa tgt gga cac aat gat gaa atg tct aga gag aat att agt gag atc aga
  L V E K C G H N D E M S R E N I S E I R
aaa gat att gtc aat ttc cat ggc gaa atg gtt ctg cta gta aac tac agt aac atc aat
  K D I V N F H G E M V L L V N Y S N I N
tac act gga tta gca aag att cta aag aag tac gac aag cga aca aga gga gga tta aga
  Y T G L A K I L K K Y D K R T R G G L R
tca cca ttt att caa aaa gtt ctt cat caa ccg ttt ttc aag act gat ctt gtt tca aga
  S P F I Q K V L H Q P F F K T D L V S R
cta gta aga gag tgg gag acg acg atg gac gcg gtg gat cca gtg aag gtg gcg gag gcg
  L V R E W E T T M D A V D P V K V A E A
```

```

gag gga tac gag aga tgt gcg gcg gtg act tcg gca gcg gcg gga gaa ggg ata ttt agg
E G Y E R C A A V T S A A A G E G I F R
aat acg gtt gcg gca tta ttg act atg aaa gag atg aga aga gga agt tcg act tac agt
N T V A A L L T M K E M R R G S S T Y S
gca ttc tca ctt ccg ccg cta aat atc tcc gat tcc gat aat gtt ctc cga tct ctt cat
A F S L P P L N I S D S D N V L R S L H
cta tct tct ccg att cct att cca taa agc ttt cta gac cag ttt gtg att aac ctc agg
L S S P I P I P - S F L D Q F V I N L R
tgc agg ctg cct atc aga agg tgg tgg ctg
C R L P I R R W W L

```

AI4.3 Example sequence data for *AtPHT5;1* in *pOPINF* from Mix2seq and ExPASy translate.

DNA_sequence

```

ggagttatccgggacctttaattcaaccaacacaatatattatagttaaataagaattattatcaaatcatttgatattaattaaa
atactatactgtaaattacattttattacaatcaaaggagatataccatggcacaccatcaccaccatcacagcagcggtctg
gaagttctgtttcagggcccgatggtggctttgggaaacttgcagcggaaacaaatcgaagaatggagtggtctattatat
caattacaaattgatgaagaagaagtgaagcaatatgctgaacaaatccaaggcggatctcaacatcctcgccatgttctca
aagatttctcgaggatgctcgatactcagattgagacaactgtcctttcatgttgaacaacaagggttcttcagggcgatt
agccaaattgagggaatctcatgatgctatacttgagcagcctgacatatcaagaatttcgagctacgtgaagcatacagag
atgttgagcagaccttctcagctcctgaaattcgttgagttgaaccattggtctgcgcaagatacttaagaaattcgaca
aaagtttgatagattcgtgattattacgtgaagaccgcgctaacccttactctcagcttcaacaagttttaagcat
gtgggtgttgagctgttgttgagcaattcccgcaatctcatgagcttcaagaaaatgaaggaagctttattcaattatga
ccaaccgttctccggctcaggatccagtgggtgaggaataaataacgcggtggacaagttaacctctcgacgaatttcc
tcaactcttgacacaacatgctcttatcatgcaagatgattgggtactccttcagaggatacaatcgaacggcttacc
tttaattcgttactcctgaatctaggaaacacattttttgtacat

```

3'5' Frame 3

```

ggag tta tcc ggg acc ttt aat tca acc caa cac aat ata tta tag tta aat aag aat tat
E L S G T F N S T Q H N I L - L N K N Y
tat caa atc att tgt ata tta att aaa ata cta tac tgt aaa tta cat ttt att tac aat
Y Q I I C I L I K I L Y C K L H F I Y N
caa agg aga tat acc atg gca cac cat cac cac cat cac agc agc ggt ctg gaa gtt ctg
Q R R Y T M A H H H H H S S G L E V L
ttt cag ggc ccg atg gtg gct ttt ggg aaa tac ttg cag cgg aaa caa atc gaa gaa tgg
F Q G P M V A F G K Y L Q R K Q I E E W
agt ggc tat tat atc aat tac aaa ttg atg aag aag aaa gtg aag caa tat gct gaa caa
S G Y Y I N Y K L M K K K V K Q Y A E Q
atc caa ggc gga tct caa cat cct cgc cat gtt ctc aaa gat ttc tcg agg atg ctc gat
I Q G G S Q H P R H V L L K D F S R M L D
act cag att gag aca act gtc ctt ttc atg ttg gaa caa caa ggg ttg ctt tca ggg cga
T Q I E T T V L F M L E Q Q G L L S G R
tta gcc aaa ttg agg gaa tct cat gat gct ata ctt gag cag cct gac ata tca aga att
L A K L R E S H D A I L E Q P D I S R I
ttc gag cta cgt gaa gca tac aga gat gtt gga cga gac ctt ctt cag ctc ctg aaa ttc
F E L R E A Y R D V G R D L L Q L L K F
gtt gag ttg aac gcc att ggt ctg cgc aag ata ctt aag aaa ttc gac aaa agg ttt gga
V E L N A I G L R K I L K K F D K R F G
tat aga ttc gct gat tat tac gtg aag acc cgc gct aat cac cct tac tct cag ctt caa
Y R F A D Y Y V K T R A N H P Y S Q L Q
caa gtt ttt aag cat gtg ggt gtt gga gct gtt gtt gga gca att tcc cgc aat ctt cat
Q V F K H V G V G A V V G A I S R N L H
gag ctt caa gaa aat gaa gga agc ttt tat tca att tat gac caa ccc gtt ctt ccg gct
E L Q E N E G S F Y S I Y D Q P V L P A
cag gat cca gtg gtt gag gca ata aat aac gcg gtg gac aag tta acc ttc tcg acg aat
Q D P V V E A I N N A V D K L T F S T N
ttc ctc aac ttc ttg gca caa cat gct ctt atc atg caa gat gat ttg gtg act cct tca
F L N F L A Q H A L I M Q D D L V T P S
gag gat aca atc gat gaa cgg tct tac cat ttt aat tcg tta ctc ctg aat cta gga aac
E D T I D E R S Y H F N S L L L N L G N
aca ttt ttt gta cat
T F F V H

```

Appendix 15 Sequence verification of cloning SPX domains from AtSPX3 and AtPHT5;1 in pGAPZA

A15.1 Example sequence data for SPX^{AtSPX3} in pGAPZA from Mix2seq and ExPASy translate.

DNA_sequence

```
CgaacacctttccccaattttggtttctcctgaccCAAagactttaatttaattta
tttgtccctatttcaatcaattgaacaactatttcgaaacgatgagatttccttcaa
tttttactgctgtttttattcgcagcatcctccgattagctgctccagtcaacacta
caacagaagatgaaacggcacaattccggctgaagctgtcatcggttactcagatt
tagaaggggatttcgatgttgctgttttgccattttccaacagcacaataacgggt
tattgtttataaataactactattgccagcattgctgctaagaagaaggggtatctc
tcgagaaaagagaggctgaagctgaattcaagtttgaaagaggattaaagaacaga
tacaagagtcgttgccggagtggcgagacaagtttcttcgttacaaggaactcaaga
atctgatctcttctccggcgccggtggaatctattttcgtcggtttgttgaacgcag
agatcgacaagtttaatgctttcttcgtcgaacaagaagaagatttcatcatccacc
acaaggagttgcaatatcggattcagagattggtagagaaatgtggacacaatgatg
aatgtctagagagaatattagtgagatcagaaaagatattgtcaatttccatggcg
aatggttctgctagtaaactacagtaacatcaattacactggattagcaaagattc
taaagaagtacgacaagcgaacaagaggaggattaagatcaccatttattcaaaaag
ttcttcatcaacttctagaacaaaaactcatctcagaagaggatctgaatagcgccg
tcgaccatcatcatcatcattgagtttgtagccttagacatgactgttcctcag
ttcaagttgggcacttacgagaggaaccgggtcttgctagattctaataagaggatg
tcagaatgccatttgccgaaagatgcaggcttcatttttgatacttttttatttg
taacctatatagga
```

5'3' Frame 3

```
cgaac acc ttt ccc caa ttt tgg ttt ctc ctg acc caa aga ctt taa att taa ttt att tgt
  N T F P Q F W F L L T Q R L - I - F I C
ccc tat ttc aat caa ttg aac aac tat ttc gaa acg atg aga ttt cct tca att ttt act
  P Y F N Q L N N Y F E T M R F P S I F T
gct gtt tta ttc gca gca tcc tcc gca tta gct gct cca gtc aac act aca aca gaa gat
  A V L F A A S S A L A A P V N T T T E D
gaa acg gca caa att ccg gct gaa gct gtc atc ggt tac tca gat tta gaa ggg gat ttc
  E T A Q I P A E A V I G Y S D L E G D F
gat gtt gct gtt ttg cca ttt tcc aac agc aca aat aac ggg tta ttg ttt ata aat act
  D V A V L P F S N S T N N G L L F I N T
act att gcc agc att gct gct aaa gaa gaa ggg gta tct ctc gag aaa aga gag gct gaa
  T I A S I A A K E E G V S L E K R E A E
gct gaa ttc aag ttt gga aag agg att aaa gaa cag ata caa gag tcg ttg ccg gag tgg
  A E F K F G K R I K E Q I Q E S L P E W
cga gac aag ttt ctt cgt tac aag gaa ctc aag aat ctg atc tct tct ccg gcg ccg gtg
```

```

R D K F L R Y K E L K N L I S S P A P V
gaa tct att ttc gtc ggt ttg ttg aac gca gag atc gac aag ttt aat gct ttc ttc gtc
E S I F V G L L N A E I D K F N A F F V
gaa caa gaa gaa gat ttc atc atc cac cac aag gag ttg caa tat cgg att cag aga ttg
E Q E E D F I I H K E L Q Y R I Q R L
gta gag aaa tgt gga cac aat gat gaa atg tct aga gag aat att agt gag atc aga aaa
V E K C G H N D E M S R E N I S E I R K
gat att gtc aat ttc cat ggc gaa atg gtt ctg cta gta aac tac agt aac atc aat tac
D I V N F H G E M V L L V N Y S N I N Y
act gga tta gca aag att cta aag aag tac gac aag cga aca aga gga gga tta aga tca
T G L A K I L K K Y D K R T R G G L R S
cca ttt att caa aaa gtt ctt cat caa ctt cta gaa caa aaa ctc atc tca gaa gag gat
P F I Q K V L H Q L L E Q K L I S E E D
ctg aat agc gcc gtc gac cat cat cat cat tga gtt tgt agc ctt aga cat gac
L N S A V D H H H H H - V C S L R H D
tgt tcc tca gtt caa gtt ggg cac tta cga gag gaa ccg gtc ttg cta gat tct aat caa
C S S V Q V G H L R E E P V L L D S N Q
gag gat gtc aga atg cca ttt gcc tga aaa gat gca ggc ttc att ttt gat act ttt tta
E D V R M P F A - K D A G F I F D T F L
ttt gta acc tat ata gga
F V T Y I G

```

Predicted size of protein with alpha factor cleaved:

Protparam = 20084.98 mw

A15.2 Example sequence data for SPX^{AtPHT5;1} in pGAPZA from Mix2seq and ExPASy translate.

DNA_sequence

```

ctcctgacccaaagactttaaatTTAATTTATTTGTCCCTATTTCAATCAATTGAAC
aactatTTTCGAAACGATGAGATTTCTTCAATTTTACTGCTGTTTTATTTCGCAGCA
tcctccgcattagctgctccagTCAACACTACAACAGAAGATGAAACGGCACAATTT
ccggctgaagctgtcatcggttactcagatttagaaggggatttcgatgTTGCTGTT
ttgccatTTTCCAACAGCACAATAACGGGTTATTGTTTTATAAATACTACTATTGCC
agcattgctgctaaagaagaaggggtatctctcgagaaaagagaggctgaagctgaa
ttcgtggctTTTGGGAAATACTTGCAGCGGAAACAAATCGAAGAATGGAGTGGCTAT
tatatcaattacaattgatgaagaagaaagtgaagcaatatgctgaacaaatccaa
ggcggatctcaacatcctcgccatgttctcaaagatttctcgaggatgctcgatact
cagattgagacaactgtcctTTTcatgTTGGAACAACAAGGGTTGCTTTcagggcga
ttagccaaattgagggaaTctcatgatgctatacttgagcagcctgacatatcaaga
atTTTCGAGCTACGTGAAGCATAcagagatgTTGGACGAGACCTTCTCagctcctg
aaattcgttgagTTGAACGCCATTGGTCTGCGCAAGATACTTAAGAAATTCGACAAA
aggTTTGGATATagattcgctgattattacgtgaagaccCGCGCTaatcacccttac
tctcagcttcaacttctagaacAAAAactcatctcagaagaggatctgaatagcGCC
gtcgaccatcatcatcatcattgagTTTGTAGCCTTAGACATGACTGTTcctca
gttcaagTTGGGCacttacgagagaaaccGGTCTTgctagatttctaatcaagaggat
gtcagaatgccattTGCCTGAAAA

```

5'3' Frame 1

```
ctc ctg acc caa aga ctt taa att taa ttt att tgt ccc tat ttc aat caa ttg aac aac
L L T Q R L - I - F I C P Y F N Q L N N
tat ttc gaa acg atg aga ttt cct tca att ttt act gct gtt tta ttc gca gca tcc tcc
Y F E T M R F P S I F T A V L F A A S S
gca tta gct gct cca gtc aac act aca aca gaa gat gaa acg gca caa att ccg gct gaa
A L A A P V N T T T E D E T A Q I P A E
gct gtc atc ggt tac tca gat tta gaa ggg gat ttc gat gtt gct gtt ttg cca ttt tcc
A V I G Y S D L E G D F D V A V L P F S
aac agc aca aat aac ggg tta ttg ttt ata aat act act att gcc agc att gct gct aaa
N S T N N G L L F I N T T I A S I A A K
gaa gaa ggg gta tct ctc gag aaa aga gag gct gaa gct gaa ttc gtg gct ttt ggg aaa
E E G V S L E K R E A E A E F V A F G K
tac ttg cag cgg aaa caa atc gaa gaa tgg agt ggc tat tat atc aat tac aaa ttg atg
Y L Q R K Q I E E W S G Y Y I N Y K L M
aag aag aaa gtg aag caa tat gct gaa caa atc caa ggc gga tct caa cat cct cgc cat
K K K V K Q Y A E Q I Q G G S Q H P R H
ggt ctc aaa gat ttc tcg agg atg ctc gat act cag att gag aca act gtc ctt ttc atg
V L K D F S R M L D T Q I E T V L F M
ttg gaa caa caa ggg ttg ctt tca ggg cga tta gcc aaa ttg agg gaa tct cat gat gct
L E Q Q G L L S G R L A K L R E S H D A
ata ctt gag cag cct gac ata tca aga att ttc gag cta cgt gaa gca tac aga gat gtt
I L E Q P D I S R I F E L R E A Y R D V
gga cga gac ctt ctt cag ctc ctg aaa ttc gtt gag ttg aac gcc att ggt ctg cgc aag
G R D L L Q L L K F V E L N A I G L R K
ata ctt aag aaa ttc gac aaa agg ttt gga tat aga ttc gct gat tat tac gtg aag acc
I L K K F D K R F G Y R F A D Y Y V K T
cgc gct aat cac cct tac tct cag ctt caa ctt cta gaa caa aaa ctc atc tca gaa gag
R A N H P Y S Q L Q L L E Q K L I S E E
gat ctg aat agc gcc gtc gac cat cat cat cat cat cat tga gtt tgt agc ctt aga cat
D L N S A V D H H H H H H - V C S L R H
gac tgt tcc tca gtt caa gtt ggg cac tta cga gag aaa ccg gtc ttg cta gat tct aat
D C S S V Q V G H L R E K P V L L D S N
caa gag gat gtc aga atg cca ttt gcc tga aaa
Q E D V R M P F A - K
```

Predicted size of protein with alpha factor cleaved:
Protparam = 21212.38 mw

Appendix 16 Purification of AtSPX1

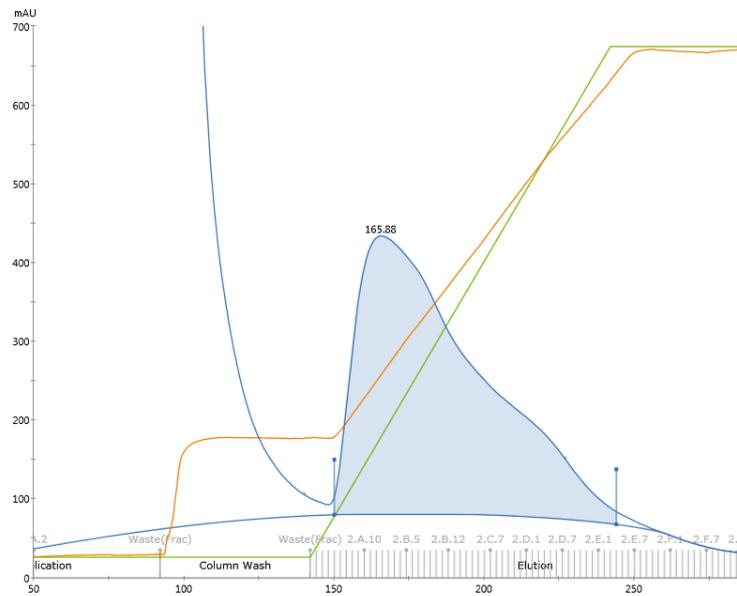


Figure A16.1 Example of UV trace data from affinity chromatography of 6xHis-tagged AtSPX1. Supernatant from lysed expression cultures were loaded onto 5 mL NiNTA column and run at 0.5 mL/min. Protein peak fractions highlighted in blue were collected for SDS-PAGE analysis.

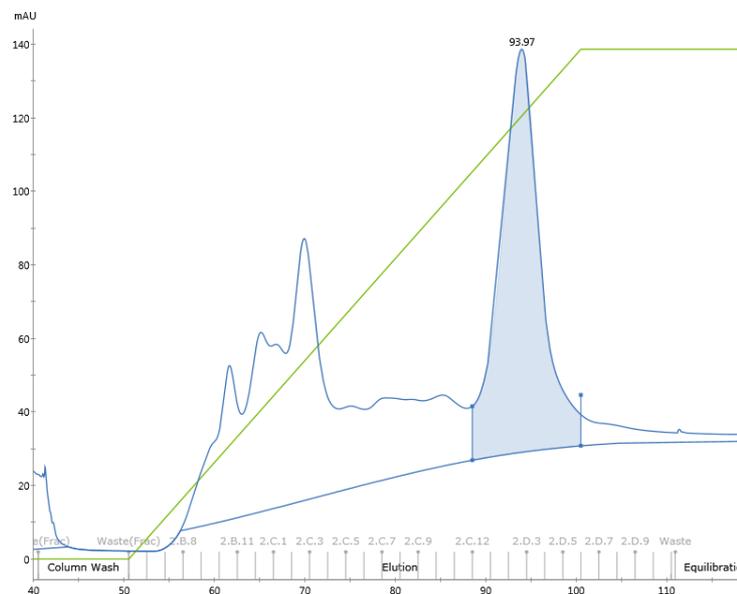


Figure A16.2 Example of UV trace data from heparin chromatography of untagged AtSPX1. Protein peak fractions highlighted in blue were collected for SDS-PAGE analysis.

Appendix 17 AtSPX1-2-FAM-InsP₅ dependence on NaCl concentration

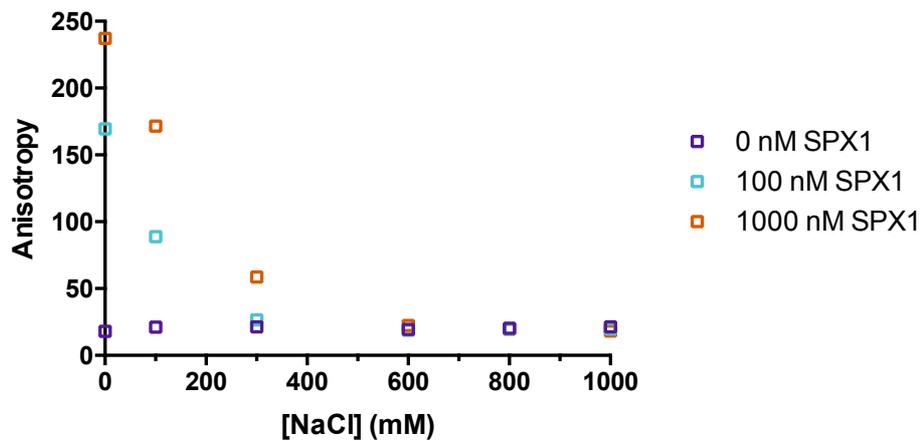


Figure A17 AtSPX1 binding 2-FAM-InsP₅ in different NaCl conditions (0-1 M), monitored by FP. Each point generated by Prism v6.0 (Graphpad) showing the mean \pm SD of fluorescence polarization data. These binding assays were performed in 20 mM HEPES pH 6.5, 1 mM MgCl₂ and incubated for 1 h before plate readings at 25 °C.

Appendix 18 Displacement of AtIPK1-2-FAM-InsP₅ by DNA

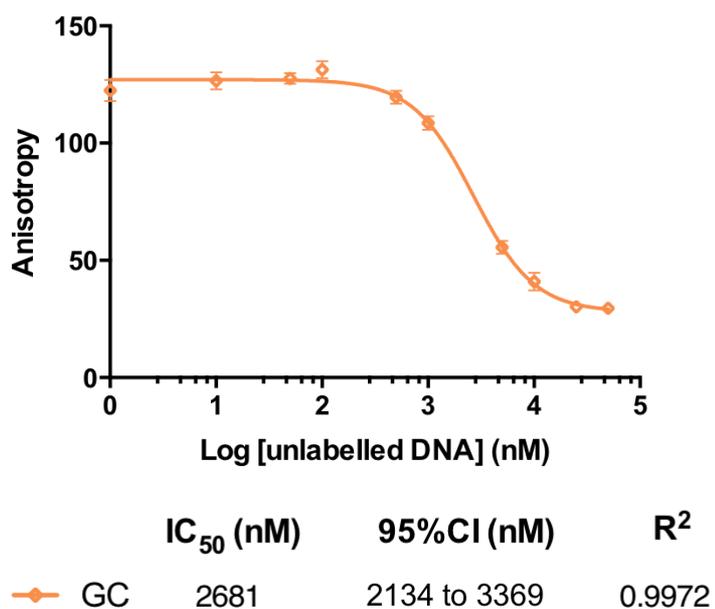


Figure A18 Displacement of AtSPXI binding of 2-FAM-InsP₅ with DNA; monitored by FP. Variable slope curves generated by Prism v6.0 (Graphpad) showing the mean \pm SD of fluorescence polarization data. These binding assays were performed in 20 mM HEPES pH 6.5, 1 mM MgCl₂, 100 mM NaCl and incubated for 1 h before plate readings at 25 °C.

Appendix 19 AtSPX1 binding different 5'-FAM-P1BS probes

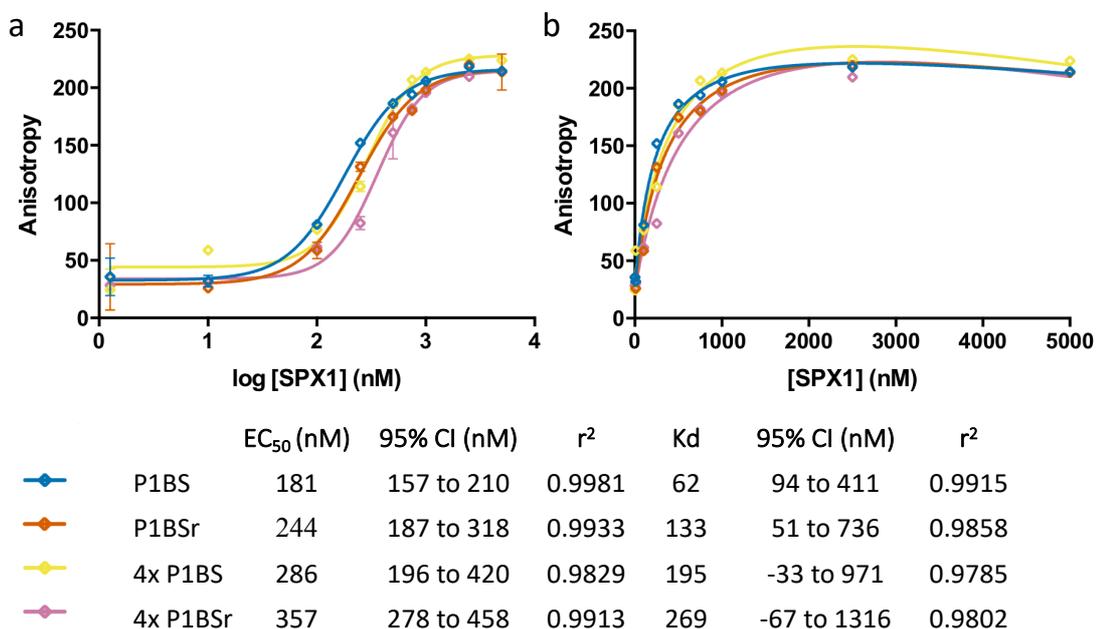


Figure A19 AtSPX1 binding FAM-P1BS in binding buffer with different additions of Pi (0-10 mM) to binding buffer (20 mM HEPES pH 6.5, 1 mM MgCl₂, 100 mM NaCl), monitored by FP. Variable slope curves generated by Prism v6.0 (Graphpad) showing the mean ± SD of anisotropy data. These binding assays were incubated for 1 h before plate readings at 25 °C.

Appendix 20 Attempted F-EMSA with AtSPX1 and FAM-dodecamer

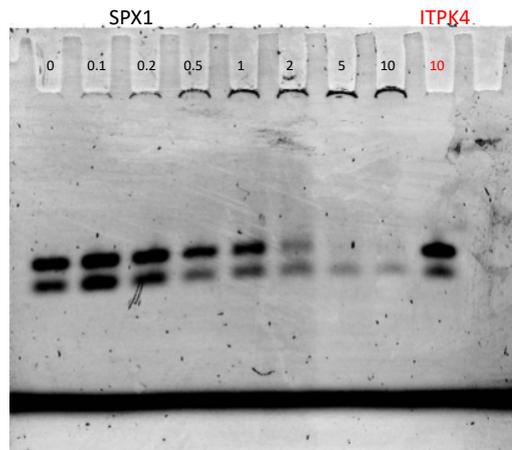


Figure A20 F-EMSA attempt with AtSPX1 and FAM-dodecamer. Using (0-10 μM) AtSPX1 and 5 nM FAM-dodecamer in binding reactions (20 mM HEPES pH 6.5, 1 mM MgCl₂, 50 mM NaCl, 0.1 mg/mL BSA, 1 mM EDTA), run on an 8% NATIVE gel and visualised by FAM detection with Typhoon FLA9.1000.

Appendix 21 AtSPX1 binding 5'-FAM-P1BS with additions of Pi

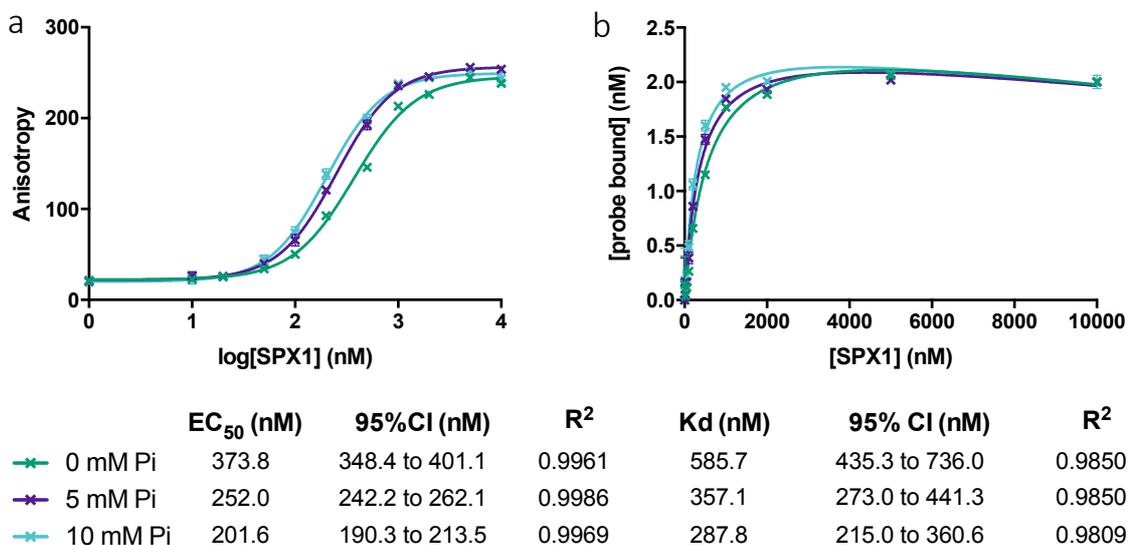


Figure A21 AtSPX1 binding 2 nM FAM-P1BS in binding buffer with different additions of Pi (0-10 mM) to binding buffer (20 mM HEPES pH 6.5, 1 mM MgCl₂, 100 mM NaCl), monitored by FP. Variable slope curves generated by Prism v6.0 (Graphpad) showing the mean ± SD of anisotropy data. These binding assays were incubated for 1 h before plate readings at 25 °C.

Appendix 22 AtSPX1 binding P1BS/2-FAM-InsP₅ with additions of Pi or pH buffer change

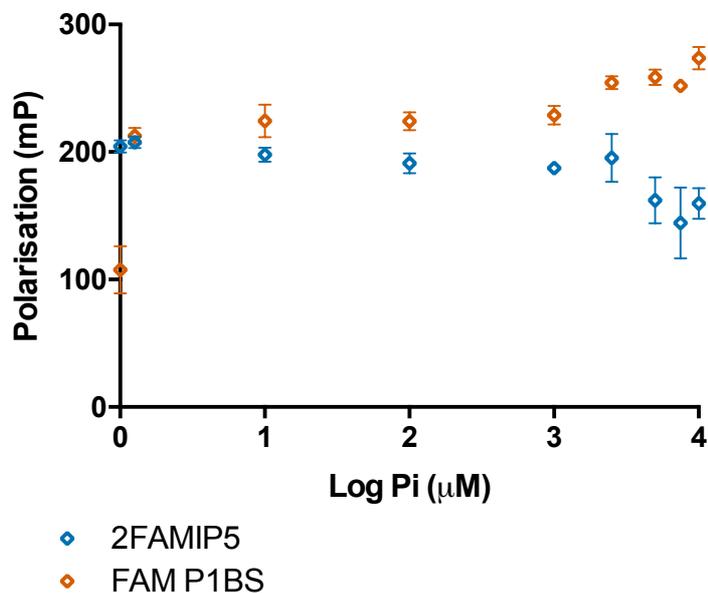


Figure A22.1 AtSPX1 binding 2 nM 2-FAM-InsP₅ or 2 nM FAM-P1BS in buffer with additions of 0 to 10 mM Pi, monitored by FP. Binding buffer contained 20 mM HEPES pH 6.5, 1 mM MgCl₂, 100 mM NaCl. Each point shows the mean \pm SD of anisotropy data.

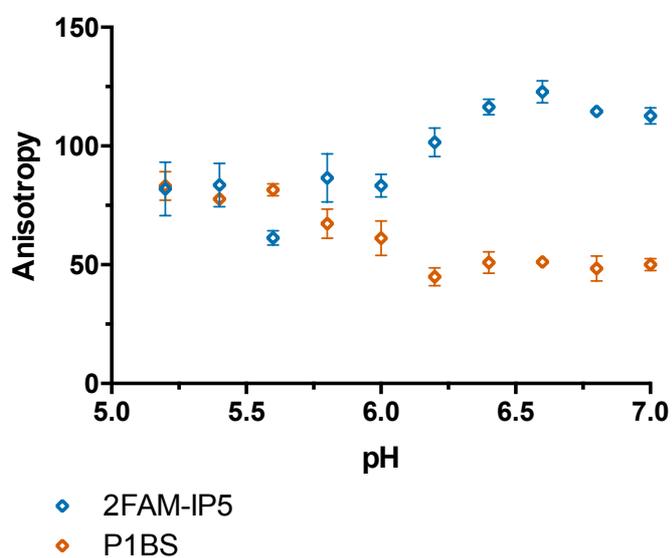


Figure A22.2 AtSPX1 binding 2 nM 2-FAM-InsP₅ or 2 nM FAM-P1BS in buffer with a range of pH concentrations, monitored by FP. Binding buffer contained 20 mM HEPES pH 5.2-7.0, 1 mM MgCl₂, 100 mM NaCl. Each point shows the mean \pm SD of anisotropy data. Readings were calibrated against free probe in the same pH buffer.

Abbreviations

5PP-InsP ₄	5-diphospho- <i>myo</i> -inositol 1,3,4,6-tetrakisphosphate
5PTases	Inositol polyphosphate 5-phosphatases
6xHis	Polyhistidine tag
ABA	Abscisic acid
ADP	Adenosine diphosphate
ASK	Arabidopsis Skp1-like
AtIPK1	Arabidopsis Inositol polyphosphate kinase 1, InsP5 2K
AtIPK2	Arabidopsis Inositol polyphosphate kinase 2, IPMK
AtITPK1	Arabidopsis Inositol tris/tetrakisphosphate kinase 1
AtITPK4	Arabidopsis Inositol tris/tetrakisphosphate kinase 4
ATP	Adenosine triphosphate
AtPHT5	Arabidopsis Phosphate transporter 5
AtSPX1	Arabidopsis SPX1
AtSPX3	Arabidopsis SPX3
AUC	Analytical ultracentrifuge
BAH1	Benzoic acid hypersensitive 1
BLAST	Basic local alignment search
C-IP	C-terminal InsP binding site
CC _{1/2}	Correlation coefficient of random half-dataset
CI	Confidence interval
COI1	Coronatine insensitive 1
CtGdel	<i>Chaetomium thermophilium</i> Glycerophosphocholine phosphodiesterase 1
CtVtc4	<i>Chaetomium thermophilium</i> Vacuole transporter chaperone 4
cv	Column volumes
DBP	DNA binding proteins
Dbp5	DEAD box protein 5
DLS	Diamond light source
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
EC50	Half maximal effective concentration
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
F-EMSA	Fluorescence-based electrophoretic mobility shift assay
FLIM	Fluorescence lifetime imaging microscopy
FP	Fluorescence polarization
FRET	Förster resonance energy transfer
GST	Glutathione-S-transferase
HEPES	4- (2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMG	High mobility group
HPLC	High performance liquid chromatography
IC50	Half maximal inhibitory concentration
IMP	Inositol monophosphatase
Ins	<i>myo</i> -inositol

InsP	Inositol phosphate
InsP1	<i>myo</i> -inositol monophosphate
InsP2	<i>myo</i> -inositol bisphosphate
InsP3	<i>myo</i> -inositol trisphosphate
InsP4	<i>myo</i> -inositol tetrakisphosphate
InsP5	<i>myo</i> -inositol pentakisphosphate
InsP6	<i>myo</i> -inositol hexakisphosphate, phytic acid
InsP7	Diphosphoinositol pentakisphosphate
InsP8	bis-diphosphoinositol tetrakisphosphate
IPK	Inositol phosphate kinase
IPK1	Inositol 1,3,4,5,6-pentakisphosphate 2-kinase, Inositol polyphosphate kinase-1
IPK2	Inositol polyphosphate multi-kinase, Inositol polyphosphate kinase-2
IPTG	Isopropyl β -D-1-thiogalactopyranoside
ITC	Isothermal calorimetry
ITPK	Inositol tris/tetraphosphate kinase
ITPK1	Inositol tetrakisphosphate 1-kinase
ITPK4	Inositol tetrakisphosphate 4-kinase
JAZ	Jasmonate ZIM domain
<i>K_d</i>	Dissociation rate at equilibrium
LB	Lysogeny broth
LBG	LB-glucose
LSC	Lysine surface cluster
MI	D- <i>myo</i> -inositol
MIK	<i>myo</i> -inositol kinase
MIPS	<i>myo</i> -inositol-3-phosphate synthase
mRNA	messenger ribonucleic acid
MT	Microscale thermophoresis
MW	Molecular weight
mwco	Molecular weight cut off
n/a	Not applicable
Ni-NTA	Nickel-nitrilotriacetic acid
NLA	Nitrogen limitation adaptation
NMR	Nuclear magnetic resonance
ns	Non-specific
nt	nucleotides
O/N	Overnight
P	Phosphate
PIBS	PHR1 binding site
PAGE	Polyacrylamide gel electrophoresis
PBC	Phosphate binding cluster
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PES	Polyethersulfone
PHO1	Phosphate 1
Pho81	Phosphatase 81
PHT5	Phosphate transporter 5

Pi	Orthophosphate, inorganic phosphate
PI4K	Phosphatidylinositol 4-kinase
PIP5K	Phosphatidylinositol-4-phosphate 5-kinase
PIS	Phosphatidylinositol synthase
PLC	Phospholipase C
PP	Pyrophosphate
PP-InsP	Inositol pyrophosphates, diphosphorylinositolphosphates
PPIP5K	Diphosphoinositol pentakisphosphate kinase
PSI	Phosphate starvation inducible
PSR	Phosphate starvation response
PtdIns	Phosphatidylinositol
PtdInsP2	Phosphatidylinositol bisphosphate
Rfree	Free residual factor
Rmerge	Residual factor on data reduction
RMSD	Root mean square deviation
rt	Room temperature
Rwork	Residual factor
SAXS	Small angle X-ray scattering
SCF	Skp1/Cdc53/F-box protein
ScIPK2	<i>Saccharomyces cerevisiae</i> inositol polyphosphate kinase-2, inositol phosphate mutikinase, IPMK
ScPho4	<i>Saccharomyces cerevisiae</i> Phosphate 1
ScVtc4	<i>Saccharomyces cerevisiae</i> Vacuole transporter chaperone 4
SD	Standard deviation
SDM	Site directed mutagenesis
SDS	Sodium dodecyl sulfate
SE	Sedimentation equilibrium
SO	Sytox orange
SPX	Sygl/pho81/xpr1
SPX-EXS	SPX-Erd1/Xpr1/Sygl
SPX-MFS	SPX-major facilitator superfamily
SPX-RING	SPX-really interesting new gene
ssDNA	Single-stranded DNA
SV	Sedimentation velocity
Sygl	Suppressor of yeast gpal
T-DNA	Transfer-DNA
TBS	Tris-buffered saline solution
TCSPC	Time correlated single-photon counting
TEMED	N, N, N', N'-tetramethylethylenediamine
TIRI	Transport inhibitor response 1
UV	Ultraviolet
VIH	VIPI homologue
VPT	Vacuolar phosphate transporter
wt	Wild-type
Xpr1	Xenotropic and polytropic retrovirus receptor 1
ZF	Zinc finger

Amino acid code:

A	Ala	Alanine
B	Asx	Asparagine or Aspartic acid
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
X	X	any codon
Y	Tyr	Tyrosine
Z	Glx	Glutamine or Glutamic acid
-	-	stop codon

References

- Adams, P. D. *et al.* (2010) 'PHENIX: A comprehensive Python-based system for macromolecular structure solution', *Acta Crystallographica Section D: Biological Crystallography*. doi: 10.1107/S0907444909052925.
- Adepoju, O. *et al.* (2019) 'Inositol Trisphosphate Kinase and Diphosphoinositol Pentakisphosphate Kinase Enzymes Constitute the Inositol Pyrophosphate Synthesis Pathway in Plants', *bioRxiv*, p. 724914. doi: 10.1101/724914.
- Afek, A. *et al.* (2014) 'Protein-DNA binding in the absence of specific base-pair recognition', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1410569111.
- Afek, A. and Lukatsky, D. B. (2012) 'Nonspecific protein-DNA binding is widespread in the yeast genome', *Biophysical Journal*. doi: 10.1016/j.bpj.2012.03.044.
- Agback, P. *et al.* (1998) 'Architecture of nonspecific protein-DNA interactions in the Sso7d-DNA complex', *Nature Structural Biology*, 5(7), pp. 579–584. doi: 10.1038/836.
- Agranoff, B. W. (1978) 'Textbook Errors-Cyclitol Confusion', *Trends in Biochemical Sciences*. ELSEVIER SCI LTD THE BOULEVARD, LANGFORD LANE, KIDLINGTON, OXFORD, OXON ..., 3(12), pp. N283–N285.
- Alberts B, Johnson A, Lewis J, *et al.* *Molecular Biology of the Cell*. 4th edition. New York: Garland Science; 2002. The Transport of Molecules between the Nucleus and the Cytosol.
Available from: <https://www.ncbi.nlm.nih.gov/books/NBK26932/>
- Alcázar-Román, A. R. *et al.* (2006) 'Inositol hexakisphosphate and Gle1 activate the DEAD-box protein Dbp5 for nuclear mRNA export.', *Nature cell biology*, 8(7), pp. 711–716. doi: 10.1038/ncb1427.
- Alcázar-Román, A. R. (2008) 'Inositol hexakisphosphate and Gle1P direct mRNA export by spatially regulating the dead-box protein Dbp5P at the nuclear pore complex'. Vanderbilt University.
- Altschul, S. F. *et al.* (1997) 'Gapped BLAST and PSI-BLAST: A new generation of protein database search programs', *Nucleic Acids Research*. doi: 10.1093/nar/25.17.3389.
- Aravind, L. and Landsman, D. (1998) 'AT-hook motifs identified in a wide variety of DNA-binding proteins', *Nucleic Acids Research*, 26(19), pp. 4413–4421. doi: 10.1093/nar/26.19.4413.
- Arkhipova, V., Guskov, A. and Slotboom, D. J. (2017) 'Analysis of the quality of

crystallographic data and the limitations of structural models', *Journal of General Physiology*. doi: 10.1085/jgp.201711852.

Arpat, A. B. *et al.* (2012) 'Functional expression of PHO1 to the Golgi and trans-Golgi network and its role in export of inorganic phosphate', *Plant Journal*. doi: 10.1111/j.1365-313X.2012.05004.x.

Baek, M. *et al.* (2017) 'GalaxyHomomer: A web server for protein homo-oligomer structure prediction from a monomer sequence or structure', *Nucleic Acids Research*. doi: 10.1093/nar/gkx246.

Bajar, B. T. *et al.* (2016) 'A guide to fluorescent protein FRET pairs', *Sensors (Switzerland)*. doi: 10.3390/sl6091488.

Balla, T. (2013) 'Phosphoinositides: Tiny Lipids With Giant Impact on Cell Regulation', *Physiological Reviews*. American Physiological Society, 93(3), pp. 1019–1137. doi: 10.1152/physrev.00028.2012.

Baños-Sanz *et al.* (2012)a 'Conformational changes in inositol 1,3,4,5,6-pentakisphosphate 2-kinase upon substrate binding: Role of N-terminal lobe and enantiomeric substrate preference', *Journal of Biological Chemistry*, 287(35), pp. 29237–29249. doi: 10.1074/jbc.M112.363671.

Baños-Sanz *et al.* (2012)b 'Expression, purification, crystallization and preliminary X-ray diffraction analysis of the apo form of InsP 5 2-K from *Arabidopsis thaliana*', *Acta Crystallographica Section F Structural Biology and Crystallization Communications*, 68(6), pp. 701–704. doi: 10.1107/S1744309112017307.

Barabote, R. D. *et al.* (2006) 'Extra domains in secondary transport carriers and channel proteins', *Biochimica et Biophysica Acta - Biomembranes*. doi: 10.1016/j.bbamem.2006.06.018.

Barnett, J. E., Rasheed, a and Corina, D. L. (1973) 'Partial reactions of D-glucose 6-phosphate-1L-myoinositol 1-phosphate cyclase.', *The Biochemical journal*, 131(1), pp. 21–30. doi: 10.1042/bj1310021.

Bauer, M. M. T. and Schnapp, G. (2007) '3.19 - Protein Production for Three-Dimensional Structural Analysis', in Taylor, J. B. and Triggler, D. J. B. T.-C. M. C. I. I. (eds). Oxford: Elsevier, pp. 411–432. doi: <https://doi.org/10.1016/B0-08-045044-X/00092-4>.

Berg, R. H. and Beachy, R. N. B. T.-M. in C. B. (2008) 'Fluorescent Protein Applications in Plants', in *Fluorescent Proteins*. Academic Press, pp. 153–177. doi: [https://doi.org/10.1016/S0091-679X\(08\)85008-X](https://doi.org/10.1016/S0091-679X(08)85008-X).

Berger, M. F. and Bulyk, M. L. (2009) 'Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of

transcription factors', *Nature Protocols*, 4(3), pp. 393–411. doi: 10.1038/nprot.2008.195.

Berman, H. M. *et al.* (2002) 'The protein data bank', *Acta Crystallographica Section D: Biological Crystallography*. doi: 10.1107/S0907444902003451.

Bernaodat, F. *et al.* (2011) 'Heterologous expression of membrane proteins: Choosing the appropriate host', *PLoS ONE*. doi: 10.1371/journal.pone.0029191.

Berridge, M. J. (1987) 'INOSITOL TRISPHOSPHATE AND DIACYLGLYCEROL: TWO INTERACTING SECOND MESSENGERS', *Annual Review of Biochemistry*. Annual Reviews, 56(1), pp. 159–193. doi: 10.1146/annurev.bi.56.070187.001111.

Berrow, N. S. *et al.* (2007) 'A versatile ligation-independent cloning method suitable for high-throughput expression screening applications', *Nucleic Acids Research*. doi: 10.1093/nar/gkm047.

Bieleski, R. L. (1973) 'Phosphate Pools, Phosphate Transport, and Phosphate Availability', *Annual Review of Plant Physiology*. doi: 10.1146/annurev.pp.24.060173.001301.

Blouin, S. *et al.* (2009) 'Functional studies of DNA-protein interactions using FRET techniques', *Methods in Molecular Biology*. doi: 10.1007/978-1-60327-015-1_28.

Blum, M. *et al.* (2021) 'The InterPro protein families and domains database: 20 years on', *Nucleic Acids Research*, 49(D1), pp. D344–D354. doi: 10.1093/nar/gkaa977.

Blumwald E., Aharon G. S. and Apse M. P. (2000) 'Sodium transport in plant cells', *Biochim Biophys Acta*, 1465: pp. 140–51. doi: 10.1016/s0005-2736(00)00135-8.

Boratyn, G. M. *et al.* (2012) 'Domain enhanced lookup time accelerated BLAST', *Biology direct*. BioMed Central, 7, p. 12. doi: 10.1186/1745-6150-7-12.

Boyer, R. F. (2002) *Concepts in biochemistry*. Brooks/Cole Pacific Grove, CA.
Brearley, C. A. and Hanke, D. E. (1996) 'Metabolic evidence for the order of addition of individual phosphate esters to the myo -inositol moiety of inositol hexakisphosphate in the duckweed *Spirodela polyrhiza* L.', *Biochemical Journal*, 314, pp. 227–233.

Brehm, M. A. *et al.* (2007) 'Intracellular localization of human Ins(1,3,4,5,6)P5 2-kinase', *Biochemical Journal*, 408(3), pp. 335–345. doi: 10.1042/BJ20070382.

Brehm, M. A. *et al.* (2013) 'A non-catalytic role for inositol 1,3,4,5,6-pentakisphosphate 2-kinase in the synthesis of ribosomal RNA', *Journal of Cell Science*, 126(2), pp. 437–444. doi: 10.1242/jcs.110031.

Buchan, D. W. A. and Jones, D. T. (2019) 'The PSIPRED Protein Analysis Workbench: 20 years on', *Nucleic Acids Research*, 47(W1), pp. W402–W407. doi: 10.1093/nar/gkz297.

Bücherl, C. A. *et al.* (2014) 'FRET-FLIM applications in plant systems', *Protoplasma*. doi: 10.1007/s00709-013-0595-7.

Buchler, N. E. and Louis, M. (2008) 'Molecular Titration and Ultrasensitivity in Regulatory Networks', *Journal of Molecular Biology*, 384(5), pp. 1106–1119. doi: <https://doi.org/10.1016/j.jmb.2008.09.079>.

Burgess, R. R. (2018) 'A brief practical review of size exclusion chromatography: Rules of thumb, limitations, and troubleshooting', *Protein Expression and Purification*, 150, pp. 81–85. doi: <https://doi.org/10.1016/j.pep.2018.05.007>.

Burra, P. V. *et al.* (2009) 'Global distribution of conformational states derived from redundant models in the PDB points to non-uniqueness of the protein structure (Proceedings of the National Academy of Sciences of the United States of America (2009) 106, 26 (10505-10510) DOI: 10.1', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0907397106.

Caddick, S. E. K. *et al.* (2007) 'A lysine accumulation phenotype of ScIpk2Delta mutant yeast is rescued by *Solanum tuberosum* inositol phosphate multikinase', *The Biochemical journal*. Portland Press Ltd., 403(3), pp. 381–389. doi: 10.1042/BJ20061772.

Caddick, S. E. K. *et al.* (2008) 'A *Solanum tuberosum* inositol phosphate kinase (StITPK1) displaying inositol phosphate–inositol phosphate and inositol phosphate–ADP phosphotransferase activities', *FEBS Letters*. John Wiley & Sons, Ltd, 582(12), pp. 1731–1737. doi: <https://doi.org/10.1016/j.febslet.2008.04.034>.

Cai, Y.-H. and Huang, H. (2012) 'Advances in the study of protein–DNA interaction', *Amino acids*. Springer, 43(3), pp. 1141–1146.

Camborde, L. *et al.* (2017) 'Detection of nucleic acid–protein interactions in plant leaves using fluorescence lifetime imaging microscopy', *Nature Protocols*. doi: 10.1038/nprot.2017.076.

Chai, C., Xie, Z. and Grotewold, E. (2011) 'SELEX (systematic evolution of ligands by exponential enrichment), as a powerful tool for deciphering the protein–DNA interaction space', *Methods in Molecular Biology*. doi: 10.1007/978-1-61779-154-3_14.

Chamberlain, P. P. *et al.* (2007) 'Integration of inositol phosphate signaling pathways via human ITPKI', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M703121200.

Chatree, S. *et al.* (2020) 'Role of Inositols and Inositol Phosphates in Energy Metabolism', *Molecules (Basel, Switzerland)*. MDPI, 25(21), p. 5079. doi: 10.3390/molecules25215079.

Chen, R. (2011) 'Bacterial expression systems for recombinant protein production: E. coli and beyond', *Biotechnology advances*, 30, pp. 1102–1107. doi: 10.1016/j.biotechadv.2011.09.013.

Chen, X., Bhadauria, V. and Ma, B. (2017) 'ChIP-Seq: A Powerful Tool for Studying Protein-DNA Interactions in Plants', in *Next-generation Sequencing and Bioinformatics for Plant Science*. doi: 10.21775/9781910190654.11.

Chorin, B. *et al.* (2020) 'ConSurf-DB: An accessible repository for the evolutionary conservation patterns of the majority of PDB proteins', *Protein Science*. doi: 10.1002/pro.3779.

Chou, Y.-H. *et al.* (2015) 'Protein-Protein Interactions Among Xyloglucan-Synthesizing Enzymes and Formation of Golgi-Localized Multiprotein Complexes', *Plant and Cell Physiology*, 56(2), pp. 255–267. doi: 10.1093/pcp/pcu161.

Cole, J. L. *et al.* (2008) 'Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium', *Methods in Cell Biology*. doi: 10.1016/S0091-679X(07)84006-4.

Collins, B., Stevens, R. C. and Page, R. (2005) 'Crystallization optimum solubility screening: Using crystallization results to identify the optimal buffer for protein crystal formation', *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*. doi: 10.1107/S1744309105035244.

Datta, R. *et al.* (2020) 'Fluorescence lifetime imaging microscopy: fundamentals and advances in instrumentation, analysis, and applications', *Journal of Biomedical Optics*. doi: 10.1117/1.jbo.25.7.071203.

Demeler, B. (2005) 'A Comprehensive Data Analysis Software Package for Analytical Ultracentrifugation Experiments', in *Modern Analytical Ultracentrifugation: Techniques and Methods*.

Desai, M. *et al.* (2014) 'Two inositol hexakisphosphate kinases drive inositol pyrophosphate synthesis in plants', *Plant Journal*. doi: 10.1111/tpj.12669.

Desfougères, Y. *et al.* (2019) 'ITPK1 mediates the lipid-independent synthesis of inositol phosphates controlled by metabolism', *Proceedings of the National Academy of Sciences of the United States of America*. 2019/11/21. National Academy of Sciences, 116(49), pp. 24551–24561. doi: 10.1073/pnas.1911431116.

Dieck, C. B., Boss, W. F. and Perera, I. Y. (2012) 'A role for phosphoinositides in regulating plant nuclear functions', *Frontiers in Plant Science*, 16;3:50. doi:

10.3389/fpls.2012.00050.

Donahue, J. L. *et al.* (2010) 'The Arabidopsis thaliana Myo-Inositol 1-Phosphate Synthase1 Gene Is Required for Myo-inositol Synthesis and Suppression of Cell Death', *The Plant Cell*, 22(3), pp. 888–903. doi: 10.1105/tpc.109.071779.

Dong, J. *et al.* (2019) 'Inositol Pyrophosphate InsP8 Acts as an Intracellular Phosphate Signal in Arabidopsis', *Molecular Plant*. doi: 10.1016/j.molp.2019.08.002.

Du, H. *et al.* (2017) 'Genome-wide identification and characterization of SPX domain-containing members and their responses to phosphate deficiency in brassica napus', *Frontiers in Plant Science*. doi: 10.3389/fpls.2017.00035.

Duan, K. *et al.* (2008) 'Characterization of a sub-family of Arabidopsis genes with the SPX domain reveals their diverse functions in plant tolerance to phosphorus starvation', *Plant Journal*. doi: 10.1111/j.1365-313X.2008.03460.x.

Eisenberg Jr, F., Bolden, A. H. and Loewus, F. A. (1964) 'Inositol formation by cyclization of glucose chain in rat testis', *Biochemical and biophysical research communications*. Elsevier, 14(5), pp. 419–424.

Emsley, P. *et al.* (2010) 'Features and development of Coot', *Acta Crystallographica Section D: Biological Crystallography*. doi: 10.1107/S09074444910007493.

Endo-Streeter, S. *et al.* (2012) 'Structural studies and protein engineering of inositol phosphate multikinase', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M112.365031.

Eubel, H., Braun, H. P. and Millar, A. H. (2005) 'Blue-native PAGE in plants: A tool in analysis of protein-protein interactions', *Plant Methods*. doi: 10.1186/1746-4811-1-11.

Felle H. H. (2005) 'pH regulation in anoxic plants', *Annals of Botany*; 96(4), pp. 519-32. doi: 10.1093/aob/mci207.

Fileppi, M. *et al.* (2010) 'Characterisation of structural genes involved in phytic acid biosynthesis in common bean (*Phaseolus vulgaris* L.)', *Molecular Breeding*. Springer, 25(3), pp. 453–470.

Fonfría-Subirós, E. *et al.* (2012) 'Crystal Structure of a Complex of DNA with One AT-Hook of HMGAI', *PLoS ONE*. Edited by E. Buratti, 7(5), p. e37120. doi: 10.1371/journal.pone.0037120.

Förster, T. (1946) 'Energiewanderung und Fluoreszenz', *Naturwissenschaften*, 33(6), pp. 166–175. doi: 10.1007/BF00585226.

Flyvbjerg H., Keatch S. A. and Dryden D. T. (2006) 'Strong physical constraints on sequence-specific target location by proteins on DNA molecules', *Nucleic Acids Res.* 12;34(9):2550-7. doi: 10.1093/nar/gkl271.

Furkert, D., Nadler-Holly, M. and Fiedler, D. (2021) 'Affinity enrichment and identification of inositol poly- and pyrophosphate interactomes', *STAR protocols*. Elsevier, 2(1), p. 100277. doi: 10.1016/j.xpro.2020.100277.

Geilfus C. M. (2018) 'Chloride: from Nutrient to Toxicant', *Plant Cell Physiol*, 1;59(5), pp. 877-886. doi: 10.1093/pcp/pcy071.

Godin, A. G. *et al.* (2011) 'Revealing protein oligomerization and densities in situ using spatial intensity distribution analysis', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1018658108.

González, B. *et al.* (2004) 'Structure of a human inositol 1,4,5-trisphosphate 3-kinase: Substrate binding reveals why it is not a phosphoinositide 3-kinase', *Molecular Cell*. doi: 10.1016/j.molcel.2004.08.004.

González, B. *et al.* (2010) 'Inositol 1,3,4,5,6-pentakisphosphate 2-kinase is a distant IPK member with a singular inositide binding site for axial 2-OH recognition.', *Proceedings of the National Academy of Sciences USA*, 107(21), pp. 9608–9613. doi: 10.1073/pnas.0912979107.

Gorrec, F. (2009) 'The MORPHEUS protein crystallization screen', *Journal of Applied Crystallography*. doi: 10.1107/S0021889809042022.

Gosein, V. *et al.* (2012) 'Inositol phosphate-induced stabilization of inositol 1,3,4,5,6- pentakisphosphate 2-kinase and its role in substrate specificity', *Protein Science*, 21(5), pp. 737–742. doi: 10.1002/pro.2049.

Gosein, V. and Miller, G. J. (2013) 'Roles of phosphate recognition in inositol 1,3,4,5,6-pentakisphosphate 2-kinase (IPK1) substrate binding and activation', *Journal of Biological Chemistry*, 288(37), pp. 26908–26913. doi: 10.1074/jbc.M113.487777.

Goswami, D., Thakker, J. N. and Dhandhukia, P. C. (2016) 'Portraying mechanics of plant growth promoting rhizobacteria (PGPR): A review', *Cogent Food & Agriculture*. doi: 10.1080/23311932.2015.1127500.

Grimm, C. *et al.* (2010) 'A crystallization screen based on alternative polymeric precipitants', *Acta Crystallographica Section D: Biological Crystallography*. doi: 10.1107/S0907444910009005.

Hamburger, D. *et al.* (2002) 'Identification and characterization of the Arabidopsis PHO1 gene involved in phosphate loading to the xylem', *Plant Cell*. doi: 10.1105/tpc.000745.

Hao, D., Ohme-Takagi, M. and Sarai, A. (1998) 'Unique Mode of GCC Box Recognition by the DNA-binding Domain of Ethylene-responsive Element-binding Factor (ERF Domain) in Plant *', *Journal of Biological Chemistry*. Elsevier, 273(41), pp. 26857–26861. doi: 10.1074/jbc.273.41.26857.

Hart, D. J. *et al.* (1999) 'The salt dependence of DNA recognition by NF- κ B p50: A detailed kinetic analysis of the effects on affinity and specificity', *Nucleic Acids Research*. doi: 10.1093/nar/27.4.1063.

Hellman, L. M. and Fried, M. G. (2007) 'Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions', *Nature Protocols*. doi: 10.1038/nprot.2007.249.

Hermans, C. *et al.* (2013) An update on magnesium homeostasis mechanisms in plants, *Metallomics*, 5(9), pp. 1170–1183. doi: 10.1039/c3mt20223b

von Hippel, P. H. and Berg, O. G. (1986) 'On the specificity of DNA-protein interactions', *Proceedings of the National Academy of Sciences*, 83(6), pp. 1608 LP – 1612. doi: 10.1073/pnas.83.6.1608.

Hollis, T. (2007) 'Crystallization of protein-DNA complexes', *Methods in Molecular Biology*. doi: 10.1385/1-59745-209-2:225.

Howlett, G. J., Minton, A. P. and Rivas, G. (2006) 'Analytical ultracentrifugation for the study of protein association and assembly', *Current Opinion in Chemical Biology*. doi: 10.1016/j.cbpa.2006.08.017.

Huang, K.-L., Ma, G.-J., Mei-Li Zhang, M.-L., Xiong, H., Wu, H., Zhao, C.-Z., *et al.* (2018) 'The ARF7 and ARF19 Transcription Factors Positively Regulate PHOSPHATE STARVATION RESPONSE1 in Arabidopsis Roots', *Plant Physiology*, 178 (1), pp. 413–427, <https://doi.org/10.1104/pp.17.01713>

Hwang, S., Guo, Z. and Kuznetsov, I. B. (2007) 'DP-Bind: A web server for sequence-based prediction of DNA-binding residues in DNA-binding proteins', *Bioinformatics*. doi: 10.1093/bioinformatics/btl672.

Jameson, D. M. and Ross, J. A. (2010) 'Fluorescence polarization/anisotropy in diagnostics and imaging', *Chemical Reviews*. doi: 10.1021/cr900267p.

Jancarik, J. and Kim, S. H. (1991) 'Sparse matrix sampling. A screening method for crystallization of proteins', *Journal of Applied Crystallography*. doi: 10.1107/S0021889891004430.

Jauch, R. *et al.* (2003) 'The Zinc Finger-Associated Domain of the *Drosophila* Transcription Factor Grauzone Is a Novel Zinc-Coordinating Protein-Protein Interaction Module', *Structure*. Elsevier, 11(11), pp. 1393–1402. doi: 10.1016/j.str.2003.09.015.

Jia, Q. *et al.* (2019) 'The Function of Inositol Phosphatases in Plant Tolerance to Abiotic Stress', *International journal of molecular sciences*. MDPI, 20(16), pp. 3999. doi: 10.3390/ijms20163999.

Jung, J. Y. *et al.* (2018) 'Control of plant phosphate homeostasis by inositol pyrophosphates and the SPX domain', *Current Opinion in Biotechnology*. doi: 10.1016/j.copbio.2017.08.012.

Kant, S., Peng, M. and Rothstein, S. J. (2011) 'Genetic regulation by NLA and microRNA827 for maintaining nitrate-dependent phosphate homeostasis in Arabidopsis', *PLoS Genetics*. doi: 10.1371/journal.pgen.1002021.

Karbalaei, M., Rezaee, S. A. and Farsiani, H. (2020) 'Pichia pastoris: A highly successful expression system for optimal synthesis of heterologous proteins', *Journal of Cellular Physiology*. doi: 10.1002/jcp.29583.

Khodakovskaya, M. *et al.* (2010) 'Increasing inositol (1,4,5)-trisphosphate metabolism affects drought tolerance, carbohydrate metabolism and phosphate-sensitive biomass increases in tomato', *Plant Biotechnology Journal*. John Wiley & Sons, Ltd, 8(2), pp. 170–183. doi: <https://doi.org/10.1111/j.1467-7652.2009.00472.x>.

Kibbe, W. A. (2007) 'OligoCalc: an online oligonucleotide properties calculator', *Nucleic acids research*. 2007/04/22. Oxford University Press, 35(Web Server issue), pp. W43–W46. doi: 10.1093/nar/gkm234.

Kikhney, A. G. and Svergun, D. I. (2015) 'A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins', *FEBS Letters*, 589(19, Part A), pp. 2570–2577. doi: <https://doi.org/10.1016/j.febslet.2015.08.027>.

Kim, S.-I. and Tai, T. H. (2011) 'Identification of genes necessary for wild-type levels of seed phytic acid in Arabidopsis thaliana using a reverse genetics approach', *Molecular Genetics and Genomics*, 286(2), pp. 119–133. doi: 10.1007/s00438-011-0631-2.

Klemm, S. L., Shipony, Z. and Greenleaf, W. J. (2019) 'Chromatin accessibility and the regulatory epigenome', *Nature Reviews Genetics*, 20(4), pp. 207–220. doi: 10.1038/s41576-018-0089-8.

Kohler, J. J. and Schepartz, A. (2001) 'Effects of nucleic acids and polyanions on dimer formation and DNA binding by bZIP and bHLHZip transcription factors', *Bioorganic & Medicinal Chemistry*, 9(9), pp. 2435–2443. doi: [https://doi.org/10.1016/S0968-0896\(01\)00221-8](https://doi.org/10.1016/S0968-0896(01)00221-8).

Krinke, O. *et al.* (2007) 'Inositol trisphosphate receptor in higher plants: is it real?', *Journal of Experimental Botany*, 58(3), pp. 361–376. doi: 10.1093/jxb/erl220.
Krishna, S. S., Majumdar, I. and Grishin, N. V. (2003) 'Structural classification of

zinc fingers', *Nucleic Acids Research*, 31(2), pp. 532–550. doi: 10.1093/nar/gkgl61.

Krogh, A. *et al.* (2001) 'Predicting transmembrane protein topology with a hidden markov model: application to complete genomes' Edited by F. Cohen', *Journal of Molecular Biology*, 305(3), pp. 567–580. doi: <https://doi.org/10.1006/jmbi.2000.4315>.

Kufareva, I. and Abagyan, R. (2012) 'Methods of protein structure comparison', *Methods in Molecular Biology*. doi: 10.1007/978-1-61779-588-6_10.

Kumar, Anuj *et al.* (2019) 'Genome-wide identification, characterization, and expression profiling of SPX gene family in wheat', *International Journal of Biological Macromolecules*. doi: 10.1016/j.ijbiomac.2019.08.105.

Kuo, H.-F. *et al.* (2014) 'Arabidopsis inositol pentakisphosphate 2-kinase, AtIPK1, is required for growth and modulates phosphate homeostasis at the transcriptional level', *The Plant Journal*, 80(3), pp. 503–515. doi: 10.1111/tpj.12650.

Kuo, H. F. *et al.* (2018) 'Arabidopsis inositol phosphate kinases IPK1 and ITPK1 constitute a metabolic pathway in maintaining phosphate homeostasis', *Plant Journal*. doi: 10.1111/tpj.13974.

Laha, D. *et al.* (2015) 'VIH2 Regulates the Synthesis of Inositol Pyrophosphate InsP8 and Jasmonate-Dependent Defenses in Arabidopsis.', *The Plant cell*, 27(4), pp. 1082–97. doi: 10.1105/tpc.114.135160.

Laha, D. *et al.* (2019) 'Arabidopsis ITPK1 and ITPK2 Have an Evolutionarily Conserved Phytic Acid Kinase Activity', *ACS Chemical Biology*. doi: 10.1021/acscchembio.9b00423.

Laha, N. P. *et al.* (2020) 'ITPK1-Dependent Inositol Polyphosphates Regulate Auxin Responses in *Arabidopsis thaliana*', *bioRxiv*, p. 2020.04.23.058487. doi: 10.1101/2020.04.23.058487.

Land E. S., Cridland C. A., Craige B., Dye A., Hildreth S. B., Helm R. F., Gillaspay G. E. and Perera I. Y. (2021) 'A Role for Inositol Pyrophosphates in the Metabolic Adaptations to Low Phosphate in *Arabidopsis*', *Metabolites*. 4;11(9):601. doi: 10.3390/metabol1090601.

Lamb, J. J., Rokke, G. and Hohmann-Marriott, M. F. (2018) 'Chlorophyll fluorescence emission spectroscopy of oxygenic organisms at 77 K', *Photosynthetica*, 56(1), pp. 105–124. doi: 10.1007/s11099-018-0791-y.

Laue, T., Shah, B., Ridgeway, T., and Pelletier, S. (1992) 'Analytical ultracentrifugation in biochemistry and polymer science', *Analytica Chimica Acta*. doi: 10.1016/0003-2670(95)90401-8.

Laue, T. M. (1995) 'Sedimentation equilibrium as thermodynamic tool', *Methods*

in *Enzymology*. doi: 10.1016/0076-6879(95)59055-2.

Lee, H.-S. *et al.* (2015) 'InsP6-sensitive variants of the Gle1 mRNA export factor rescue growth and fertility defects of the *ipkl* low-phytic-acid mutation in *Arabidopsis*.' *The Plant cell*, 27(2), pp. 417–31. doi: 10.1105/tpc.114.132134.

Lemtiri-Chlieh, F. *et al.* (2003) 'Inositol hexakisphosphate mobilizes an endomembrane store of calcium in guard cells', *Proceedings of the National Academy of Sciences of the United States of America*. 2003/08/11. National Academy of Sciences, 100(17), pp. 10091–10095. doi: 10.1073/pnas.1133289100.

Lemtiri-Chlieh, F., MacRobbie, E. A. and Brearley, C. A. (2000) 'Inositol hexakisphosphate is a physiological signal regulating the K⁺-inward rectifying conductance in guard cells', *Proceedings of the National Academy of Sciences of the United States of America*. The National Academy of Sciences, 97(15), pp. 8687–8692. doi: 10.1073/pnas.140217497.

Lin, J. *et al.* (2012) 'Physical organization of DNA by multiple non-specific DNA-binding modes of integration host factor (IHF)', *PloS one*. 2012/11/14. Public Library of Science, 7(11), pp. e49885–e49885. doi: 10.1371/journal.pone.0049885.

Lin, W. Y., Huang, T. K. and Chiou, T. J. (2013) 'NITROGEN LIMITATION ADAPTATION, a target of MicroRNA827, Mediates degradation of plasma membrane-localized phosphate transporters to maintain phosphate homeostasis in *Arabidopsis*', *Plant Cell*. doi: 10.1105/tpc.113.116012.

Liu, G. *et al.* (2015) 'Inositol hexaphosphate suppresses growth and induces apoptosis in HT-29 colorectal cancer cells in culture: PI3K/Akt pathway as a potential target', *International Journal of Clinical and Experimental Pathology*.

Liu, H. and Naismith, J. H. (2008) 'An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol', *BMC Biotechnology*, 8(1), p. 91. doi: 10.1186/1472-6750-8-91.

Liu, J. *et al.* (2015) 'A vacuolar phosphate transporter essential for phosphate homeostasis in *Arabidopsis*', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1514598112.

Liu, T. Y. *et al.* (2016) 'Identification of plant vacuolar transporters mediating phosphate storage', *Nature Communications*. doi: 10.1038/ncomms11095.

Liu, Y.-L. *et al.* (2011) 'Functional Role of Dimerization of Human Peptidylarginine Deiminase 4 (PAD4)', *PLOS ONE*. Public Library of Science, 6(6), p. e21314. Available at: <https://doi.org/10.1371/journal.pone.0021314>.

Loewus, F. A. and Murthy, P. P. N. (2000) 'myo-Inositol metabolism in plants', *Plant Science*, 150(1), pp. 1–19. doi: 10.1016/S0168-9452(99)00150-8.

Loewus, M. W. and Loewus, F. A. (1982) 'myo-Inositol-1-Phosphatase from the Pollen of *Lilium longiflorum* Thunb', *Plant physiology*, 70(3), pp. 765–770. doi: 10.1104/pp.70.3.765.

Lohman, T. M. and von Hippel, P. H. (1986) 'Kinetics of Protein-Nucleic Acid Interactions: Use of Salt Effects to Probe Mechanisms of Interactio', *Critical Reviews in Biochemistry*. Taylor & Francis, 19(3), pp. 191–245. doi: 10.3109/10409238609084656.

López-Arredondo, D. L. *et al.* (2014) 'Phosphate Nutrition: Improving Low-Phosphate Tolerance in Crops', *Annual Review of Plant Biology*. Annual Reviews, 65(1), pp. 95–123. doi: 10.1146/annurev-arplant-050213-035949.

Luan, M. *et al.* (2019) 'Vacuolar phosphate transporters contribute to systemic phosphate homeostasis vital for reproductive development in arabidopsis l[open]', *Plant Physiology*. doi: 10.1104/pp.18.01424.

Luscombe, N. M., Laskowski, R. A. and Thornton, J. M. (2001) 'Amino acid-base interactions: A three-dimensional analysis of protein-DNA interactions at an atomic level', *Nucleic Acids Research*. doi: 10.1093/nar/29.13.2860.

Luscombe, N. M. and Thornton, J. M. (2002) 'Protein–DNA Interactions: Amino Acid Conservation and the Effects of Mutations on Binding Specificity', *Journal of Molecular Biology*, 320(5), pp. 991–1009. doi: [https://doi.org/10.1016/S0022-2836\(02\)00571-5](https://doi.org/10.1016/S0022-2836(02)00571-5).

Macbeth, M. R. *et al.* (2005) 'Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing', *Science (New York, N.Y.)*, 309(5740), pp. 1534–1539. doi: 10.1126/science.1113150.

Madeira, F. *et al.* (2019) 'The EMBL-EBI search and sequence analysis tools APIs in 2019', *Nucleic acids research*. doi: 10.1093/nar/gkz268.

Malabanan, M. M. and Blind, R. D. (2016) 'Inositol polyphosphate multikinase (IPMK) in transcriptional regulation and nuclear inositide metabolism', *Biochemical Society Transactions*, 44(1), pp. 279–285. doi: 10.1042/BST20150225.

Marianayagam, N. J., Sunde, M. and Matthews, J. M. (2004) 'The power of two: Protein dimerization in biology', *Trends in Biochemical Sciences*. doi: 10.1016/j.tibs.2004.09.006.

McPherson, A. *et al.* (1996) 'Incorporation of impurities into macromolecular crystals', *Journal of Crystal Growth*. doi: 10.1016/0022-0248(96)00364-8.

Mertens, N., Remaut, E. and Fiers, W. (1995) 'Tight Transcriptional Control Mechanism Ensures Stable High-Level Expression from T7 Promoter-Based Expression Plasmids', *Bio/Technology*, 13(2), pp. 175–179. doi: 10.1038/nbt0295-175.

Miyakawa, T. and Tanokura, M. (2017) 'Structural basis for the regulation of phytohormone receptors', *Bioscience, Biotechnology, and Biochemistry*, 81(7), pp. 1261–1273. doi: 10.1080/09168451.2017.1313696.

Mizuguchi, K. and Ahmad, S. (2014) 'Conformational changes in DNA-binding proteins: Relationships with precomplex features and contributions to specificity and stability', *Proteins: Structure, Function, and Bioinformatics*. John Wiley & Sons, Ltd, 82(5), pp. 841–857. doi: <https://doi.org/10.1002/prot.24462>.

Moll, J. R., Acharya A., Gal J., Mir A. A. and Vinson C. (2002) 'Magnesium is required for specific DNA binding of the CREB B-ZIP domain', *Nucleic Acids Res.* 30(5), pp. 1240-1246. doi: 10.1093/nar/30.5.1240. Erratum in: *Nucleic Acids Res* 2002 May 1;30(9):2096. Gal Jozset [corrected to Gal Jozsef]. PMID: 11861917; PMCID: PMC101231

Monserrate, J. P. and York, J. D. (2010) 'Inositol phosphate synthesis and the nuclear processes they affect', *Current Opinion in Cell Biology*. Elsevier Ltd, 22(3), pp. 365–373. doi: 10.1016/j.ceb.2010.03.006.

Morgunova, E. and Taipale, J. (2017) 'Structural perspective of cooperative transcription factor binding', *Current Opinion in Structural Biology*. doi: 10.1016/j.sbi.2017.03.006.

Mosblech, A. *et al.* (2011) 'Jasmonic acid perception by COII involves inositol polyphosphates in *Arabidopsis thaliana*', *The Plant Journal*, 65(6), pp. 949–957. doi: 10.1111/j.1365-313X.2011.04480.x.

Müller, S. *et al.* (2013) 'Quantification of Förster resonance energy transfer by monitoring sensitized emission in living plant cells', *Frontiers in Plant Science*, p. 413. Available at: <https://www.frontiersin.org/article/10.3389/fpls.2013.00413>.

Murashige, T. and Skoog, F. (1962) 'A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures', *Physiologia Plantarum*. doi: 10.1111/j.1399-3054.1962.tb08052.x.

Murphy, A. M. *et al.* (2008) 'A role for inositol hexakisphosphate in the maintenance of basal resistance to plant pathogens', *Plant Journal*, 56(4), pp. 638–652. doi: 10.1111/j.1365-313X.2008.03629.x.

Najet, A. *et al.* (2000) 'ArgRII, a Component of the ArgR-McmI Complex Involved in the Control of Arginine Metabolism in *Saccharomyces cerevisiae*, Is the Sensor of Arginine', *Molecular and Cellular Biology*. American Society for Microbiology, 20(6), pp. 2087–2097. doi: 10.1128/MCB.20.6.2087-2097.2000.

Newman, J. *et al.* (2005) 'Towards rationalization of crystallization screening for small- To medium-sized academic laboratories: The PACT/JCSG+ strategy', *Acta Crystallographica Section D: Biological Crystallography*. doi:

10.1107/S0907444905024984.

'Numbering of atoms in myo-inositol. Recommendations 1988. Nomenclature Committee of the International Union of Biochemistry' (1989) *The Biochemical journal*, 258(1), pp. 1–2. Available at: <https://pubmed.ncbi.nlm.nih.gov/2930499>.

Ockenden, I. *et al.* (2004) 'Characterization of the storage of phosphorus, inositol phosphate and cations in grain tissues of four barley (*Hordeum vulgare* L.) low phytic acid genotypes', *Plant Science*. Elsevier, 167(5), pp. 1131–1142.

Odom, A. R. *et al.* (2000) 'A role for nuclear inositol 1,4,5-trisphosphate kinase in transcriptional control.', *Science (New York, N.Y.)*, 287(5460), pp. 2026–9. doi: 10.1126/science.287.5460.2026.

Osorio, M. B. *et al.* (2019) 'SPX4 Acts on PHR1-dependent and -independent regulation of shoot phosphorus status in arabidopsis', *Plant Physiology*. doi: 10.1104/pp.18.00594.

Otegui, M. S., Capp, R. and Staehelin, L. A. (2002) 'Developing Seeds of Arabidopsis Store Different Minerals in Two Types of Vacuoles and in the Endoplasmic Reticulum', *The Plant Cell*, 14(6), pp. 1311–1327. doi: 10.1105/tpc.010486.

Owczarzy, R., Moreira B. G., You, Y., Behlke, M. A. and Walder J. A. (2008) 'Predicting Stability of DNA Duplexes in Solutions Containing Magnesium and Monovalent Cations', *Biochemistry* 47(19), pp. 5336–5353. doi: 10.1021/bi702363u

Parthasarathy, R. and Eisenberg Jr, F. (1986) 'The inositol phospholipids: a stereochemical view of biological activity', *The Biochemical journal*, 235(2), pp. 313–322. doi: 10.1042/bj2350313.

Paulus, H. and Kennedy, E. (1960) 'The Enzymatic Synthesis of Inositol Monophosphatide', *Journal of Biological Chemistry*, 235, pp. 1303–1311. doi: 10.1016/S0021-9258(18)69403-7.

Paz, I. *et al.* (2016) 'BindUP: a web server for non-homology-based prediction of DNA and RNA binding proteins', *Nucleic acids research*. doi: 10.1093/nar/gkw454.

Peng, M. *et al.* (2007) 'A mutation in NLA, which encodes a RING-type ubiquitin ligase, disrupts the adaptability of Arabidopsis to nitrogen limitation', *Plant Journal*. doi: 10.1111/j.1365-3113.2007.03050.x.

Perrin, F. (1926) 'Polarisation de la lumière de fluorescence. Vie moyenne des molécules dans l'état excité', *Journal de Physique et le Radium*. Société Française de Physique, 7(12), pp. 390–401.

Phillippy, B. Q., Ullah, A. H. and Ehrlich, K. C. (1994) 'Purification and some

properties of inositol 1,3,4,5,6-Pentakisphosphate 2-kinase from immature soybean seeds.', *Journal of Biological Chemistry*, 269(45), pp. 28393–28399. doi: [https://doi.org/10.1016/S0021-9258\(18\)46940-2](https://doi.org/10.1016/S0021-9258(18)46940-2).

Pilchova, I., Klacanova, K., Tatarkova, Z., Kaplan, P. and Racay, P. (2017) 'The Involvement of Mg²⁺ in Regulation of Cellular and Mitochondrial Functions', *Oxidative Medicine and Cellular Longevity*, doi: <https://doi.org/10.1155/2017/6797460>

Pilu, R. *et al.* (2003) 'Phenotypic, genetic and molecular characterization of a maize low phytic acid mutant (lpa241)', *Theoretical and Applied Genetics*. Springer, 107(6), pp. 980–987.

Pratt, J. *et al.* (2009) 'Phosphate (Pi) starvation effect on the cytosolic Pi concentration and Pi exchanges across the tonoplast in plant cells: An in vivo 31P-nuclear magnetic resonance study using methylphosphonate as a Pi analog', *Plant Physiology*. doi: 10.1104/pp.109.144626.

Prigge, M. J., Platre, M., Kadakia, N., Zhang, Y., Greenham, K., Szutu, W., Pandey, B. K., Bhosale, R. A., Bennett, M. J., Busch, W. and Estelle, M. (2020) 'Genetic analysis of the Arabidopsis TIR1/AFB auxin receptors reveals both overlapping and specialized functions', *Elife*, 18(9):e54740. doi: 10.7554/eLife.54740.

Privalov, P. L., Dragan, A. I. and Crane-Robinson, C. (2011) 'Interpreting protein/DNA interactions: Distinguishing specific from non-specific and electrostatic from non-electrostatic components', *Nucleic Acids Research*. doi: 10.1093/nar/gkq984.

Puga, M. I. *et al.* (2014) 'SPX1 is a phosphate-dependent inhibitor of Phosphate Starvation Response 1 in Arabidopsis', *Proceedings of the National Academy of Sciences of the United States of America*, 111(41). doi: 10.1073/pnas.1404654111.

Puga, M. I. *et al.* (2017) 'Novel signals in the regulation of Pi starvation responses in plants: facts and promises', *Current Opinion in Plant Biology*. doi: 10.1016/j.pbi.2017.05.007.

Purohit, A. (2018) 'Electrostatic Interactions at the Dimer Interface Stabilize the E. coli β sliding Clamp', *Biophysical Journal*. doi: 10.1016/j.bpj.2017.11.1230.

Qi, W. (2016) *Characterization of SPX exclusive family members in plant Pi sensing and regulation*. University of Leeds. Available at: <https://etheses.whiterose.ac.uk/15758/>.

Qi, W. *et al.* (2017) 'AtSPX1 affects the AtPHR1–DNA-binding equilibrium by binding monomeric AtPHR1 in solution', *Biochemical Journal*. doi: 10.1042/BCJ20170522.

Qiu, D. *et al.* (2020) 'Analysis of inositol phosphate metabolism by capillary

electrophoresis electrospray ionization mass spectrometry', *Nature communications*. Nature Publishing Group UK, 11(1), p. 6035. doi: 10.1038/s41467-020-19928-x.

Raghothama, K. G. (1999) 'Phosphate acquisition', *Annual Review of Plant Physiology and Plant Molecular Biology*, 50, pp. 665–693.

Raicu, V. and Singh, D. R. (2013) 'FRET spectrometry: A new tool for the determination of protein quaternary structure in living cells', *Biophysical Journal*. doi: 10.1016/j.bpj.2013.09.015.

Ramirez-Garcés, D. *et al.* (2016) 'CRN13 candidate effectors from plant and animal eukaryotic pathogens are DNA-binding proteins which trigger host DNA damage response', *New Phytologist*. doi: 10.1111/nph.13774.

Regvar, M. *et al.* (2011) 'New insights into globoids of protein storage vacuoles in wheat aleurone using synchrotron soft X-ray microscopy', *Journal of Experimental Botany*. Oxford University Press, 62(11), pp. 3929–3939.

Richardson, A. E. and Simpson, R. J. (2011) 'Soil microorganisms mediating phosphorus availability', *Plant Physiology*. doi: 10.1104/pp.111.175448.

Riechmann, J. L. *et al.* (2000) 'Arabidopsis Transcription Factors: Genome-Wide Comparative Analysis Among Eukaryotes', *Science*. American Association for the Advancement of Science, 290(5499), pp. 2105–2110. doi: 10.1126/science.290.5499.2105.

Ried, M. K. *et al.* (2021) 'Inositol pyrophosphates promote the interaction of SPX domains with the coiled-coil motif of PHR transcription factors to regulate plant phosphate homeostasis', *Nature Communications*. doi: 10.1038/s41467-020-20681-4.

Riemer, E. *et al.* (2021) 'ITPK1 is an InsP(6)/ADP phosphotransferase that controls phosphate signaling in Arabidopsis', *Molecular plant*. 2021/07/15. Oxford University Press, 14(11), pp. 1864–1880. doi: 10.1016/j.molp.2021.07.011.

Roshchina, V. V (2012) 'Vital Autofluorescence: Application to the Study of Plant Living Cells', *International Journal of Spectroscopy*. Edited by J. Zhang. Hindawi Publishing Corporation, 2012, p. 124672. doi: 10.1155/2012/124672.

Rubio, V. *et al.* (2001) 'A conserved MYB transcription factor involved in phosphate starvation signaling both in vascular plants and in unicellular algae', *Genes and Development*. doi: 10.1101/gad.204401.

Russo Krauss, I. *et al.* (2013) 'An overview of biological macromolecule crystallization', *International journal of molecular sciences*, 14(6), p. 11643–11691. doi: 10.3390/ijms140611643.

Saha, A., Wittmeyer, J. and Cairns, B. R. (2006) 'Chromatin remodelling: the

industrial revolution of DNA around histones', *Nature Reviews Molecular Cell Biology*, 7(6), pp. 437–447. doi: 10.1038/nrml945.

Sainsbury, F., Thuenemann, E. C. and Lomonosoff, G. P. (2009) 'pEAQ : versatile expression vectors for easy and quick transient expression of heterologous proteins in plants', *Plant Biotechnology Journal*, 7, pp. 682–693. doi: 10.1111/j.1467-7652.2009.00434.x.

Schindelin, J. *et al.* (2012) 'Fiji: An open-source platform for biological-image analysis', *Nature Methods*. doi: 10.1038/nmeth.2019.

Schroeder, J. I. *et al.* (2001) 'GUARD CELL SIGNAL TRANSDUCTION', *Annual Review of Plant Physiology and Plant Molecular Biology*. Annual Reviews, 52(1), pp. 627–658. doi: 10.1146/annurev.arplant.52.1.627.

Seeds, A. M. *et al.* (2007) 'Roles for inositol polyphosphate kinases in the regulation of nuclear processes and developmental biology', *Advances in Enzyme Regulation*, 47, pp. 10–25. doi: 10.1016/j.advenzreg.2006.12.019.

Seo, A. *et al.* (2013) 'Using UV-absorbance of intrinsic dithiothreitol (DTT) during RP-HPLC as a measure of experimental redox potential in vitro', *Analytical and Bioanalytical Chemistry*. doi: 10.1007/s00216-013-7063-2.

Serdar, C. C. *et al.* (2021) 'Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies', *Biochemia medica*. 2020/12/15. Croatian Society of Medical Biochemistry and Laboratory Medicine, 31(1), p. 10502. doi: 10.11613/BM.2021.010502.

Sheard, L. B. *et al.* (2010) 'Jasmonate perception by inositol-phosphate-potentiated COII-JAZ co-receptor', *Nature*. 2010/10/06, 468(7322), pp. 400–405. doi: 10.1038/nature09430.

Shears, S. B. (2015) 'Inositol pyrophosphates: Why so many phosphates?', *Advances in Biological Regulation*, 57, pp. 203–216. doi: 10.1016/j.jbior.2014.09.015.

Sherman, W. R., Stewart, M. A. and Zinbo, M. (1969) 'Mass Spectrometric Study on the Mechanism of d-Glucose 6-Phosphate-l-myo-Inositol 1-Phosphate Cyclase', *J. Biol. Chem.*, 244(20), pp. 5703–5708. Available at: <http://www.jbc.org/cgi/content/long/244/20/5703> (Accessed: 5 October 2016).

Sivasakthi, V., Anbarasu, A. and Ramaiah, S. (2013) ' π - π Interactions in Structural Stability: Role in RNA Binding Proteins', *Cell Biochemistry and Biophysics*. doi: 10.1007/s12013-013-9573-0.

Sloan, J. *et al.* (2020) 'Structural basis for the complex DNA binding behavior of the plant stem cell regulator WUSCHEL', *Nature Communications*. doi: 10.1038/s41467-020-16024-y.

Smith, A. P. *et al.* (2010) 'Histone H2A.Z Regulates the Expression of Several Classes of Phosphate Starvation Response Genes But Not as a Transcriptional Activator', *Plant Physiology*, 152(1), pp. 217–225. doi: 10.1104/pp.109.145532.

Steiner, S. and Pfannschmidt, T. (2009) 'Fluorescence-based Electrophoretic Mobility Shift Assay in the Analysis of DNA-binding Proteins', *Methods in molecular biology (Clifton, N.J.)*, 479, pp. 273–289. doi: 10.1007/978-1-59745-289-2_18.

Stephens, L. and Irvine, R. F. (1990) 'Stepwise phosphorylation of myo-inositol leading to myo-inositol hexakisphosphate in Dictyostelium', *Nature*, 346, pp. 580–583.

Stevenson-Paulik *et al.* (2005) 'Generation of phytate-free seeds in Arabidopsis through disruption of inositol polyphosphate kinases.', *Proceedings of the National Academy of Sciences of the United States of America*, 102(35), pp. 12612–12617. doi: 10.1073/pnas.0504172102.

Stevenson-Paulik, J., Odom, A. R. and York, J. D. (2002) 'Molecular and biochemical characterization of two plant inositol polyphosphate 6-/3-/5-kinases.', *The Journal of biological chemistry*, 277(45), pp. 42711–42718. doi: 10.1074/jbc.M209112200.

Štros, M., Launholt, D. and Grasser, K. D. (2007) 'The HMG-box: A versatile protein domain occurring in a wide variety of DNA-binding proteins', *Cellular and Molecular Life Sciences*. doi: 10.1007/s00018-007-7162-3.

Sunami, T. and Kono, H. (2013) 'Local Conformational Changes in the DNA Interfaces of Proteins', *PLOS ONE*. Public Library of Science, 8(2), p. e56080. Available at: <https://doi.org/10.1371/journal.pone.0056080>.

Svedberg, T. and Pedersen, K. O. (1940) 'The ultracentrifuge.', *The Ultracentrifuge*. Oxford: Clarendon Press.

Sweetman, D. *et al.* (2006) 'Characterization of an Arabidopsis inositol 1,3,4,5,6-pentakisphosphate 2-kinase (AtIPK1)', *Biochemical Journal*, 394(1), pp. 95–103. doi: 10.1042/BJ20051331.

Sweetman, D. *et al.* (2007) 'Arabidopsis thaliana inositol 1,3,4-trisphosphate 5/6-kinase 4 (AtITPK4) is an outlier to a family of ATP-grasp fold proteins from Arabidopsis', *FEBS Letters*. John Wiley & Sons, Ltd, 581(22), pp. 4165–4171. doi: <https://doi.org/10.1016/j.febslet.2007.07.046>.

Szabó, Á. *et al.* (2018) 'The Effect of Fluorophore Conjugation on Antibody Affinity and the Photophysical Properties of Dyes', *Biophysical Journal*. doi: 10.1016/j.bpj.2017.12.011.

Tan, X. *et al.* (2007) 'Mechanism of auxin perception by the TIR1 ubiquitin ligase', *Nature*, 446(7136), pp. 640–645. doi: 10.1038/nature05731.

Thomas, M. P., Mills, S. J. and Potter, B. V. L. (2016) 'The "Other" Inositols and Their Phosphates: Synthesis, Biology, and Medicine (with Recent Advances in myo-Inositol Chemistry)', *Angewandte Chemie International Edition*. John Wiley & Sons, Ltd, 55(5), pp. 1614–1650. doi: <https://doi.org/10.1002/anie.201502227>.

Thompson, M. and Woodbury, N. W. (2001) 'Thermodynamics of specific and nonspecific DNA binding by two DNA-binding domains conjugated to fluorescent probes', *Biophysical Journal*. doi: 10.1016/S0006-3495(01)75830-4.

Tjong, H. and Zhou, H. X. (2007) 'DISPLAR: An accurate method for predicting DNA-binding sites on protein surfaces', *Nucleic Acids Research*. doi: 10.1093/nar/gkm008.

Tsui, M. M. and York, J. D. (2010) 'Roles of inositol phosphates and inositol pyrophosphates in development, cell signaling and nuclear processes', *Advances in Enzyme Regulation*, 50(1), pp. 324–337. doi: 10.1016/j.advenzreg.2009.12.002.

Tzeng, S.-R. and Kalodimos, C. G. (2012) 'Protein activity regulation by conformational entropy', *Nature*, 488(7410), pp. 236–240. doi: 10.1038/nature11271.

Valouev, A. *et al.* (2011) 'Determinants of nucleosome organization in primary human cells', *Nature*. doi: 10.1038/nature10002.

Verbsky, J. W. *et al.* (2002) 'The synthesis of inositol hexakisphosphate: Characterization of human inositol 1,3,4,5,6-pentakisphosphate 2-kinase', *Journal of Biological Chemistry*, 277(35), pp. 31857–31862. doi: 10.1074/jbc.M205682200.

Wagner, S. *et al.* (2006) 'Rationalizing membrane protein overexpression', *Trends in Biotechnology*. doi: 10.1016/j.tibtech.2006.06.008.

Wagner, S. *et al.* (2008) 'Tuning Escherichia coli for membrane protein overexpression', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0804090105.

Wang, L. *et al.* (2021) 'Loss of two families of SPX domain-containing proteins required for vacuolar polyphosphate accumulation coincides with the transition to phosphate storage in green plants', *Molecular Plant*, 14(5), pp. 838–846. doi: <https://doi.org/10.1016/j.molp.2021.01.015>.

Wang, P., Snijders, R., Kohlen, W., Liu, J., Bisseling, T., Limpens, E. (2021) 'Medicago SPX1 and SPX3 regulate phosphate homeostasis, mycorrhizal colonization, and arbuscule degradation', *The Plant Cell*, 33(11), pp 3470–3486. doi: <https://doi.org/10.1093/plcell/koab206>.

Wang, Y. *et al.* (2004) 'Structure and expression profile of the Arabidopsis PHO1 gene family indicates a broad role in inorganic phosphate homeostasis', *Plant Physiology*. doi: 10.1104/pp.103.037945.

Wang, Z., Kuo, H.-F. and Chiou, T.-J. (2021) 'Intracellular phosphate sensing and regulation of phosphate transport systems in plants', *Plant Physiology*. doi: 10.1093/plphys/kiab343.

Watson, P. J. *et al.* (2016) 'Insights into the activation mechanism of class I HDAC complexes by inositol phosphates', *Nature communications*. Nature Publishing Group, 7, p. 11262. doi: 10.1038/ncomms11262.

Watts, A. (1993) 'Crystallization of nucleic acids and proteins. A practical approach', *FEBS Letters*. doi: 10.1016/0014-5793(93)80565-c.

Weber, G. (1952) 'Polarization of the fluorescence of macromolecules. I. Theory and experimental method.', *The Biochemical journal*. doi: 10.1042/bj0510145.

Wei, M. *et al.* (2007) 'High-throughput determination of mode of inhibition in lead identification and optimization', *Journal of Biomolecular Screening*. doi: 10.1177/1087057106296679.

Welner, D. H. *et al.* (2012) 'DNA binding by the plant-specific NAC transcription factors in crystal and solution: a firm link to WRKY and GCM transcription factors', *Biochemical Journal*, 444(3), pp. 395–404. doi: 10.1042/BJ20111742.

Whitfield, H. (2013) *Structural and biochemical studies of plant inositol phosphate kinases*. University of East Anglia. Available at: <https://ueaeprints.uea.ac.uk/id/eprint/48345/>.

Whitfield, H. *et al.* (2018) 'A Fluorescent Probe Identifies Active Site Ligands of Inositol Pentakisphosphate 2-Kinase', *Journal of Medicinal Chemistry*. doi: 10.1021/acs.jmedchem.8b01022.

Whitfield, H. *et al.* (2020) 'An ATP-responsive metabolic cassette comprised of inositol tris/tetrakisphosphate kinase 1 (ITPK1) and inositol pentakisphosphate 2-kinase (IPK1) buffers diphosphoinositol phosphate levels', *Biochemical Journal*, 477(14), pp. 2621–2638. doi: 10.1042/BCJ20200423.

Whitfield H., Hemmings A. M., Mills S. J., Baker K., White G., Rushworth S., Riley A. M., Potter B. V. L and Brearley C. A. (2021) 'Allosteric Site on SHIP2 Identified Through Fluorescent Ligand Screening and Crystallography: A Potential New Target for Intervention', *Journal of Medicinal Chemistry*. 8;64(7):3813-3826. doi: 10.1021/acs.jmedchem.0c01944.

Wild, R. *et al.* (2016) 'Control of eukaryotic phosphate homeostasis by inositol polyphosphate sensor domains', *Science*. doi: 10.1126/science.aad9858.

Wilfinger, W. W., Mackey, K. and Chomczynski, P. (1997) 'Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity', *BioTechniques*. doi: 10.2144/97223st01.

Williams, S. P., Gillaspay, G. E. and Perera, I. Y. (2015) 'Biosynthesis and possible functions of inositol pyrophosphates in plants.', *Frontiers in Plant Science*, 6(February), p. 67. doi: 10.3389/fpls.2015.00067.

Wingfield, P. T. (2015) 'Overview of the purification of recombinant proteins', *Current protocols in protein science*, 80, pp. 6.1.1-6.1.35. doi: 10.1002/0471140864.ps0601s80.

Wingfield, P. T. (2016) 'Protein precipitation using ammonium sulfate', *Current Protocols in Protein Science*. doi: 10.1002/0471140864.psa03fs84.

Withers, J., Yao, J., Mecey, C., Howe, G. A., Melotto, M. and He, S. Y. (2012) 'Transcription factor-dependent nuclear localization of a transcriptional repressor in jasmonate hormone signaling', *Proc Natl Acad Sci*, 4;109(49), pp. 20148-53. doi: 10.1073/pnas.1210054109.

Wu, P. *et al.* (2013) 'Improvement of phosphorus efficiency in rice on the basis of understanding phosphate signaling and homeostasis', *Current Opinion in Plant Biology*. doi: 10.1016/j.pbi.2013.03.002.

Xia, H.-J. (2003) 'Arabidopsis Inositol Polyphosphate 6- β -Kinase Is a Nuclear Protein That Complements a Yeast Mutant Lacking a Functional ArgR-Mcml Transcription Complex', *THE PLANT CELL ONLINE*, 15(2), pp. 449-463. doi: 10.1105/tpc.006676.

Xia, Y. *et al.* (2015) 'New insights into the QuikChange™ process guide the use of Phusion DNA polymerase for site-directed mutagenesis', *Nucleic Acids Research*, 43(2), p. e12. doi: 10.1093/nar/gkull89.

Xu, L. *et al.* (2002) 'The SCF(COII) ubiquitin-ligase complexes are required for jasmonate response in Arabidopsis', *The Plant cell*. American Society of Plant Biologists, 14(8), pp. 1919-1935. doi: 10.1105/tpc.003368.

Yan, X. *et al.* (2002) 'Evaluation of different nucleic acid stains for sensitive double-stranded DNA analysis with capillary electrophoretic separation', *Journal of Chromatography A*, 943(2), pp. 275-285. doi: [https://doi.org/10.1016/S0021-9673\(01\)01452-2](https://doi.org/10.1016/S0021-9673(01)01452-2).

Yan, Y. *et al.* (2017) 'HDOCK: A web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy', *Nucleic Acids Research*. doi: 10.1093/nar/gkx407.

Yan, Y. *et al.* (2020) 'The HDOCK server for integrated protein-protein docking',

Nature Protocols, 15(5), pp. 1829–1852. doi: 10.1038/s41596-020-0312-x.

Yang, S. (2014) 'Methods for SAXS-Based Structure Determination of Biomolecular Complexes', *Advanced Materials*. John Wiley & Sons, Ltd, 26(46), pp. 7902–7910. doi: <https://doi.org/10.1002/adma.201304475>.

Yang, S. Y. *et al.* (2017) 'Role of vacuoles in phosphorus storage and remobilization', *Journal of Experimental Botany*. doi: 10.1093/jxb/erw481.

Yao, Z., Tian, J. and Liao, H. (2014) 'Comparative characterization of GmSPX members reveals that GmSPX3 is involved in phosphate homeostasis in soybean', *Annals of Botany*. doi: 10.1093/aob/mcu147.

Yesudhas, D. *et al.* (2017) 'Proteins recognizing DNA: Structural uniqueness and versatility of DNA-binding domains in stem cell transcription factors', *Genes*. doi: 10.3390/genes8080192.

York, J. D. *et al.* (1999) 'A phospholipase C-dependent inositol polyphosphate kinase pathway required for efficient messenger RNA export.', *Science (New York, N.Y.)*, 285(5424), pp. 96–100. doi: 10.1126/science.285.5424.96.

York, J. D. (2006) 'Regulation of nuclear processes by inositol polyphosphates', *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1761(5–6), pp. 552–559. doi: 10.1016/j.bbalip.2006.04.014.

Zabet, N. R. and Adryan, B. (2013) 'The effects of transcription factor competition on gene regulation', *Frontiers in Genetics*. doi: 10.3389/fgene.2013.00197.

Zhao, H. *et al.* (2013) 'Overview of current methods in sedimentation velocity and sedimentation equilibrium analytical ultracentrifugation', *Current protocols in protein science*. Wiley Online Library, 71(1), pp. 12–20.

Zhou, J., Hu, Q., Xiao, X. *et al.* (2021) 'Mechanism of phosphate sensing and signaling revealed by rice SPX1-PHR2 complex structure', *Nat Commun* 12, 7040. <https://doi.org/10.1038/s41467-021-27391-5>

Zhou, X. *et al.* (2005) 'The dark side of green fluorescent protein', *New Phytologist*. John Wiley & Sons, Ltd, 168(2), pp. 313–322. doi: <https://doi.org/10.1111/j.1469-8137.2005.01489.x>.

Zhu, J. *et al.* (2019) 'Two bifunctional inositol pyrophosphate kinases/phosphatases control plant phosphate homeostasis', *eLife*. doi: 10.7554/eLife.43582.

Zhu, J. Y., Sun, Y. and Wang, Z. Y. (2012) 'Genome-wide identification of transcription factor-binding sites in plants using chromatin immunoprecipitation followed by microarray (ChIP-chip) or sequencing (ChIP-seq)', *Methods in Molecular Biology*. doi: 10.1007/978-1-61779-809-2_14.

Zoonens, M. and Miroux, B. (2010) 'Expression of membrane proteins at the Escherichia coli membrane for structural studies.', *Methods in molecular biology (Clifton, N.J.)*. doi: 10.1007/978-1-60761-344-2_4.

Zulauf, M. and D'Arcy, A. (1992) 'Light scattering of proteins as a criterion for crystallization', *Journal of Crystal Growth*. doi: 10.1016/0022-0248(92)90232-8.