Structural Studies of *Arabidopsis thaliana* Inositol Polyphosphate Multi-Kinase

by

Stuart Tamotsu Endo-Streeter

Department of Biochemistry
Duke University

Date: ______________________

Approved:

John D. York, Ph.D., Supervisor

Arno L. Greenleaf, Ph.D.

Christian R.H. Raetz, M.D., Ph.D.

Matthew Redinbo, Ph.D.

David C. Richardson, Ph.D.

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry in the Graduate School of Duke University

2009
ABSTRACT

Structural Studies of *Arabidopsis thaliana* Inositol Polyphosphate Multi-Kinase

by

Stuart Tamotsu Endo-Streeter

Department of Biochemistry
Duke University

Date: __________________________
Approved: __________________________

John D. York, Ph.D., Supervisor

Arno L. Greenleaf, Ph.D.

Christian R.H. Raetz, Ph.D.

Matthew Redinbo, Ph.D.

David C. Richardson, Ph.D.

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry in the Graduate School of Duke University

2009
Abstract

Inositol Polyphosphate Multi-Kinase (IPMK, also known as ArgRIII, Arg82, and IPK2) is a central component of the inositol signaling system, catalyzing the phosphorylation of at least four different inositol polyphosphate species in vivo with in vitro activity observed for three more. Each of these IP species is sterically unique and the phosphorylation target varies between the 6’-, 3’-, or 5’-hydroxyls, classifying IPMK as a 6/3/5-kinase. The products of IPMK have been linked to multiple processes including cell cycle regulation, transcriptional control, telomere length regulation, mRNA export and various phenotypes including mouse embryonic and fly larvae development, and stress responses in plants and yeast. Linking specific IP species and cellular processes has been complicated by the inability to distinguish between the different effects of the various IP species generated by IPMK. Deletion of IPMK affects the IP populations of all its various substrates and products and therefore the role of a single IP species cannot be tracked. The goals of this work were to address the question of substrate selectivity and develop new tools to probe inositol signaling in vivo through a combination of structural, enzymatic, and genomic techniques.

The structure of Arabidopsis thaliana IPMK is reported at 2.9Å resolution and in conjunction with a new model of inositol selectivity has been used to design constructs with altered substrate profiles. In vitro and in vivo experiments have confirmed that IPMK identifies substrate inositol polyphosphate species through a recognition surface
that requires phosphate groups occupy specific pockets and rejects those with axial phosphate groups in specific regions. *In vivo* experiments have linked specific inositol polyphosphate species to nitrogen metabolism and temperature sensitivity in yeast and established the potential for these constructs to be used to probe signaling in other organisms.
Dedication

This work is especially dedicated to three people. My father, Donald J. Streeter, who first encouraged me to follow my interests in science and always supported me through science fairs and projects, and most of all, by always making time to answer my questions and encouraging me to find those answers he did not know. To Robert (Bob) Madar, ret., who taught the Field and Advanced Field Biology courses at Crescent Valley High School. There I had my first opportunities to truly engage in academic research in conjunction with Oregon State University, my eventual undergraduate alma mater, which set me on a fortunate path of opportunities to reach where I am today. And finally, to P. Andrew Karplus, Ph.D, Chair of the Department Biochemistry and Biophysics, Oregon State University. As an undergraduate he gave me the opportunity to work in his lab and introduced me to the world of protein structure that I have enjoyed and been intrigued by ever since.
Table of Contents

Abstract ................................................................................................................................. iv

List of Tables ...................................................................................................................... x

List of Figures ..................................................................................................................... xi

List of Abbreviations .......................................................................................................... xiii

Acknowledgements ............................................................................................................ xvii

Introduction ......................................................................................................................... 1
  Inositol Signaling .................................................................................................................. 1
  Inositol Polyphosphate Multi-Kinase .................................................................................. 4
  Questions .............................................................................................................................. 7
  Objectives of this work ........................................................................................................ 8
  Approaching Inositol Signaling with an Integrated Methodology ..................................... 9

Chapter 1. From Gene to Density; Solving the Structure of *Arabidopsis thaliana* Inositol Polyphosphate Multi-Kinase α ................................................................. 11
  Introduction ........................................................................................................................ 11
  First Constructs and Crystals ............................................................................................. 11
  Encountering Twinning ...................................................................................................... 18
  Switching to a new tag, linker, and construct .................................................................. 20
  Solving the Phase Problem ............................................................................................... 30
  Building and Refinement .................................................................................................. 32
  Summary ............................................................................................................................ 34
Experimental Methods........................................................................................................35

Chapter 2. The Structure of Arabidopsis thaliana Inositol Polyphosphate Multi-Kinase-α and a New Substrate Selectivity Theory ................................................................. 40

The Structure of Arabidopsis thaliana IPMK ................................................................. 40
Introduction .......................................................................................................................... 40
Structural Description ....................................................................................................... 42
Similarities and Differences to Yeast IPMK and IP3K..................................................... 48
Initial Substrate Selectivity Implications ................................................................. 51

Inositol and IPMK ........................................................................................................... 51
Introduction .......................................................................................................................... 51
The Structural Richness of Inositol ................................................................................. 52
Previous Selectivity Theories ........................................................................................... 53
An Inositol Substrate Motif ............................................................................................ 56

Inositol Binding Modes in IPMK .................................................................................... 61
The inositol binding pockets of AtIpmk ......................................................................... 62
A 1'-, 4'-phosphate binding pocket is not supported ................................................... 65
Why AtIpmk is not a 1/4-kinase and specificity from steric exclusion ....................... 68
Inositol polyphosphate substrate recognition by AtIpmk ............................................. 73

Summary............................................................................................................................. 73

Experimental Methods..................................................................................................... 76

Chapter 3. Redesigning IPMK........................................................................................... 77
Introduction.......................................................................................................................... 77
Rational Design and the Inositol Binding Site ............................................................... 78
Redesigning the IP Binding Site of IPMK with KiNG..................................................... 82
Testing the Motif Hypothesis ................................................................. 91

Mutant production of IP₄ and identification of the K121W AtIpmk product .......... 91

Kinetic parameters of wild-type, K117W, and K121W AtIpmk .................................. 95

Summary ........................................................................................................ 99

Experimental Methods ............................................................................... 100

Chapter 4. Probing Inositol Signaling In-vivo .................................................. 105

Introduction ................................................................................................. 105

Metabolomics .............................................................................................. 107

Phenotypic Investigations ........................................................................... 113

Summary ........................................................................................................ 118

Experimental Methods ............................................................................... 120

Conclusions and Future Directions ............................................................... 123

Conclusions ................................................................................................. 123

Future Directions ........................................................................................ 131

Appendix ....................................................................................................... 134

References .................................................................................................... 136

Biography ..................................................................................................... 143
List of Tables

Table 2-1: Crystallographic Data, Phasing, and Refinement Statistics ....................... 41
Table 2-2: Hydrogen Bonds at helix $\alpha 8$ interface .................................................. 46
Table 3-1: Kinetics ............................................................................................................. 98
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-1</td>
<td>Inositol Stereochemistry</td>
<td>2</td>
</tr>
<tr>
<td>I-2</td>
<td>Partial Signaling Map of Inositol Polyphosphates</td>
<td>3</td>
</tr>
<tr>
<td>1-1</td>
<td>IPMK Expression Screening</td>
<td>15</td>
</tr>
<tr>
<td>1-2</td>
<td>Initial <em>At</em>IpMk crystals</td>
<td>18</td>
</tr>
<tr>
<td>1-3</td>
<td>A twinned crystal with two lattices</td>
<td>20</td>
</tr>
<tr>
<td>1-4</td>
<td>The pGEX-KG cloning site</td>
<td>24</td>
</tr>
<tr>
<td>1-5</td>
<td>pET-15b and pGEX-NB cloning site</td>
<td>25</td>
</tr>
<tr>
<td>1-6</td>
<td>Final Protein Samples</td>
<td>27</td>
</tr>
<tr>
<td>1-7</td>
<td>MALDI-ToF Spectrum of Sel-Met and Native <em>At</em>IpMk</td>
<td>29</td>
</tr>
<tr>
<td>1-8</td>
<td>Final 16-286 <em>At</em>IpMk crystals</td>
<td>29</td>
</tr>
<tr>
<td>1-9</td>
<td>Experimental density maps</td>
<td>32</td>
</tr>
<tr>
<td>2-1</td>
<td>Structure of <em>At</em>IpMk</td>
<td>43</td>
</tr>
<tr>
<td>2-2</td>
<td>Structure-based sequence alignment of <em>At</em>IpMk and <em>Sc</em>IpMk and <em>Hs</em>Ip3k</td>
<td>44</td>
</tr>
<tr>
<td>2-3</td>
<td>Helix α8 loop and interface</td>
<td>45</td>
</tr>
<tr>
<td>2-4</td>
<td>Symmetry mate crystal packing</td>
<td>47</td>
</tr>
<tr>
<td>2-5</td>
<td>Structural alignment of IP kinases</td>
<td>48</td>
</tr>
<tr>
<td>2-6</td>
<td>Active sites of <em>At</em>IpMk and <em>Hs</em>Ip3k with IP3</td>
<td>49</td>
</tr>
<tr>
<td>2-7</td>
<td>Yeast poly-aspartate insert and structural comparison</td>
<td>50</td>
</tr>
<tr>
<td>2-8</td>
<td>Comparison of equatorial hydroxyl positions in conserved 1'-4' motif</td>
<td>55</td>
</tr>
<tr>
<td>2-9</td>
<td>Substrate and non-substrate inositol polyphosphate species</td>
<td>58</td>
</tr>
<tr>
<td>2-10</td>
<td>Inositol polyphosphate species and proposed binding motif</td>
<td>59</td>
</tr>
</tbody>
</table>
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-ptase</td>
<td>inositol polyphosphate 5-phosphatase</td>
</tr>
<tr>
<td>β-Me</td>
<td>b-mercaptoethanol</td>
</tr>
<tr>
<td>Δ16, 25</td>
<td>Atlpmk with first 16 or 25 residues removed</td>
</tr>
<tr>
<td>σ</td>
<td>sigma</td>
</tr>
<tr>
<td>Amp</td>
<td>ampicillin</td>
</tr>
<tr>
<td>AMP</td>
<td>adenosine 5’-monophosphate</td>
</tr>
<tr>
<td>AMP-PnP</td>
<td>adenosine 5’-(β, γ-imido) triphosphate</td>
</tr>
<tr>
<td>APS</td>
<td>Advanced Photon Source</td>
</tr>
<tr>
<td>A. thaliana, At</td>
<td><em>Arabidopsis thaliana</em></td>
</tr>
<tr>
<td>ASU</td>
<td>asymmetric unit</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine 5’-triphosphate</td>
</tr>
<tr>
<td>C-terminal</td>
<td>carboxy-terminal</td>
</tr>
<tr>
<td>CCD</td>
<td>charge coupled diode</td>
</tr>
<tr>
<td>CCP4</td>
<td>Collaborative Computational Project No. 4</td>
</tr>
<tr>
<td>CNS</td>
<td>Crystallography and NMR Systems</td>
</tr>
<tr>
<td>COOT</td>
<td>Crystallographic Object-Oriented Toolkit</td>
</tr>
<tr>
<td>cpm</td>
<td>count per minute, measure of radioactivity</td>
</tr>
<tr>
<td>cv</td>
<td>column volume</td>
</tr>
<tr>
<td>D. melanogaster, Dm</td>
<td><em>Drosophila melanogaster</em></td>
</tr>
</tbody>
</table>
DNA  deoxyribonucleic acid

E. coli  *Escherichia coli*

EDTA  ethylenediaminetetraacetate

$g$  gravitational force; $9.8\text{m/s}^2$

$\gamma$  gamma

GST  glutathione-S-transferase

*H. sapiens, Hs*  *Homo sapiens sapiens*

HEPES  4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HPLC  high-performance liquid chromatography

$I/\sigma$  signal to noise ratio

IP  inositol polyphosphate

$IP_3$  inositol-1,4,5-triskisphosphate

$IP_4$  inositol tetrakisphosphate (when identity unknown)

$IP_5$  inositol-1,3,4,5,6-pentakisphosphate

$IP_6$  inositol-1,2,3,4,5,6-hexakisphosphate, also known as phytate

IP3K  Inositol-1,4,5-trisphosphate 3-Kinase

IPMK  Inositol Polyphosphate Multi-Kinase

IPK1  Inositol Polyphosphate Kinase 1

IPK2  Inositol Polyphosphate Kinase 2 (aka, IPMK)

IPTG  isopropyl $\beta$-D-1-thiogalactopyranoside

kDa  kiloDalton

KiNG  Kinemage, Next Generation
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LB</td>
<td>Luria-Bertani media (10g tryptone, 5g yeast extract, 10g NaCl per liter)</td>
</tr>
<tr>
<td>M9 minimal media</td>
<td>minimal media made with M9 salts (200ml 5x M9 salts ((64g Na₂HPO₄, 15g KH₂PO₄, 2.5 NaCl, 5g NH₄Cl)/L), 2ml of 1M MgSO₄, 20ml of 20% filter sterilized glucose, 100μl 1M CaCl₂, 100μl 0.5% thiamine and water up to 1L)</td>
</tr>
<tr>
<td>MAD</td>
<td>multi-wavelength anomalous dispersion</td>
</tr>
<tr>
<td>MALDI-TOF</td>
<td>matrix-assisted laser desorption ionization - time of flight</td>
</tr>
<tr>
<td>MCS</td>
<td>multiple cloning site</td>
</tr>
<tr>
<td>MR</td>
<td>molecular replacement</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry</td>
</tr>
<tr>
<td>N-terminal</td>
<td>amino-terminal</td>
</tr>
<tr>
<td>NCS</td>
<td>non-crystallographic symmetry</td>
</tr>
<tr>
<td>Ni-NTA</td>
<td>nickel nitrilotriacetic acid resin</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>PP-IP</td>
<td>pyrophosphate inositol polyphosphate</td>
</tr>
<tr>
<td>rmsd</td>
<td>root-mean-squared deviation</td>
</tr>
<tr>
<td>mRNA</td>
<td>ribose nucleic acid, messenger</td>
</tr>
<tr>
<td>R. norvegicus, Rn</td>
<td>Rattus norvegicus</td>
</tr>
<tr>
<td>rpm</td>
<td>revolutions per minute</td>
</tr>
<tr>
<td><strong>S. cerevisiae, Sc</strong></td>
<td><em>Saccharomyces cerevisiae</em></td>
</tr>
<tr>
<td>---------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>SAD</td>
<td>single-wavelength anomalous dispersion</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>SEC</td>
<td>size exclusion chromatography</td>
</tr>
<tr>
<td>Sel-Met</td>
<td>seleno-methionine</td>
</tr>
<tr>
<td>TLC</td>
<td>thin-layer chromatography</td>
</tr>
<tr>
<td>TLS</td>
<td>translation-libration-screw refinement</td>
</tr>
<tr>
<td>Tris</td>
<td>tris (hydroxymethyl)-aminomethane</td>
</tr>
<tr>
<td>$V_m$</td>
<td>Matthew’s coefficient specific volume</td>
</tr>
<tr>
<td>XDS</td>
<td>X-ray Detector Software</td>
</tr>
</tbody>
</table>
Acknowledgements

My thanks and gratitude to my mentor in this work, John D. York, Ph.D. He has been a constant source of inspiration and motivation. I arrived in his lab with a background in protein crystallography, but little else. I leave with much more through his dedication to more than just performing experiments, but to expanding ones knowledge and understanding of the greater story.

I also thank my committee members for their time and input: Arno Greenleaf, Ph.D., Christian Raetz, M.D, Ph.D., Matthew Redinbo, Ph.D, and David Richardson, Ph.D.

My thanks to my colleagues and companions at Duke. I thank all of my fellow members of the York lab, especially D. Eric Dollins, Ph.D, with whom I often discussed structure ideas and theories, Ace Hatch and Trang Pham, who provided valuable information and assistance regarding yeast, and Peter Fridy and Jim Otto, Ph.D., for all their time and assistance with the HPLC experiments. In the Richardson Lab I would especially thank Jeremy Block, with whom the first steps of the inositol binding motif were brainstormed one day on a chalk board, and Vincent Chen, who was always willing to help when I pushed the KiNG program to its limits.

Finally, I thank my family for their support. My father Donald J. Streeter and mother Linda H. Endo, my sister Andrea Bouche, and my wife Pingping Song.
Introduction

Inositol Signaling

D-\textit{myo}-Inositol is a simple molecule that sits at the center of the incredibly complex inositol signaling system unique to eukaryotes (Irvine, 2005; Irvine and Schell, 2001; Majerus et al., 1999; Xia and Yang, 2005; York, 2006; York et al., 2001). A six-carbon cyclic alcohol, its ability to regulate such a complex system derives from the steric richness of the phosphorylated species. Each hydroxyl about the carbon ring is in the equatorial configuration save for the 2’-hydroxyl, which is axial. The ramification of such a minor structural difference is that any substitution or addition to any position results in a stERICally unique species (Fig. I.1). Simple alterations, such as the addition of a phosphate moiety, give rise to a potential library of \(2^6\) inositol polyphosphate (IP) species, rising to \(3^6\) if pyrophosphorylation is considered. Each sterically unique species has the potential for different interactions with different molecules. Not surprisingly, nature has taken advantage of this flexibility.
Figure I-1: Inositol Stereochemistry.

A. D-myoinositol. B. Enantiomeric species. Inositol-1-monokisphosphate (right) and Inositol-3-monokisphosphate (left) are enantiomers and cannot be superimposed upon each other.

Inositol is probably best known for its involvement in the regulation of calcium signaling. Inositol-1,4,5-triskisphosphate (IP$_3$) is cleaved from phosphatidylinositol-4,5-bisphosphate (PIP$_2$) by Phospholipase-C and subsequently binds to the IP$_3$-receptor, triggering release of Ca$^{2+}$ from intracellular stores (Cui et al., 2004; Schulz and Krause, 2004; Tisi et al., 2004; Wagner et al., 2004; Wagner et al., 2003; Zhu et al., 1999). Other systems regulated by the flux of different IP species include immune cell development (Jayaraman and Marks, 1997), endoplasmic reticulum membrane-localized signaling
pathways (Jesch et al., 2006), metabolism (Bechet et al., 1970; Delforge et al., 1975; Messenguy, 1976), stress response (Dubois et al., 2002; Xiong et al., 2001), and transcription and translation control (Odom et al., 2000; Saiardi et al., 2000; York et al., 1999; York et al., 2005). In addition to the processes regulated by the soluble inositol polyphosphate species, multiple signaling roles are played by the phosphatidylinositol lipids (Cocco et al., 2004), which are not addressed in this work.

---

**Figure I-2: Partial Signaling Map of Inositol Polyphosphates**

Several of the principal kinase pathways in inositol polyphosphate signaling starting with the release of I(1,4,5)P3 from PIP2 by Plc and moving through multiple phosphorylation steps. Kinases, phosphorylation steps, and related processes are similarly colored. IPMK purple arrows for the IP3 → IP4 → IP5 steps for visibility.

Despite nearly fifty years of investigation, many details of inositol signaling remain unknown. Often only the involvement of an IP species is known, the actual mechanism of regulation remaining unclear. Meanwhile advances in techniques are leading to the discovery of new IP species in vivo, for which roles have not yet even been proposed. In yeast, the model system for many eukaryotic investigations, the major inositol polyphosphate population is inositol-1,2,3,4,5,6-hexakisphosphate (IP6) (York et
al., 1999). Other IP species such as IP₃, inositol-1,4,5,6-tetrakisphosphate, and inositol-1,3,4,5,6-pentakisphosphate are not observed. Yet clearly they must exist at some point along the synthesis route to IP₆. When genes involved in the synthesis of various IP species are disrupted the precursor species can be detected in vivo, proving their existence. These other species likely exist in a transitory or low-level state under normal conditions, with population increases only under specific circumstances as follows their function as temporary signals (Fujii and York, 2005; Seeds et al., 2005; Stevenson-Paulik et al., 2002). More recently new knowledge has come to light over the synthesis of pyrophosphorylated inositols, the PP-IP species (Fridy et al., 2007).

For the existence of so many IP species a large number of modifying enzymes must also exist, in this case inositol kinases and phosphatases. Indeed, multiple IP kinases and phosphatases have been discovered and linked to various different systems (Majerus et al., 1999; Shears, 2004; York, 2006), partially defining a large web of interactions through which cellular signals are passed. Inositol polyphosphate multi-kinase plays a central role in this signaling web.

**Inositol Polyphosphate Multi-Kinase**

Inositol polyphosphate multi-kinase was not originally discovered as kinase. It had in fact been identified in 1970 as a component in arginine metabolism. Five arginine anabolic genes are repressed and two catabolic genes are induced by the ARGRI, ARGRII, and ARGRIII (IPMK) proteins (Bechet et al., 1970). In the presence of arginine the ARGR proteins bind to the ARG3 and ARG5,6 and CAR1 and CAR2 promoters as measured by gel shift assays (Messenguy et al., 1991). This is clearly part
of a arginine-sensing and feedback mechanism. When arginine is plentiful, the cell turns off production of proteins used to generate arginine and activates the production of proteins that breakdown arginine. A loss of any of the ARGR proteins results in constitutive anabolic and no catabolic gene expression, as the complexes cannot form to repress the first or activate the second. Subsequently the ARGR proteins were found to form an ARGR-MCM1 complex, in which the general transcription factor MCM1 completed the complex and was required for functionality (El Bakkoury et al., 2000; Messenguy and Dubois, 1993). MCM1 is a promiscuous transcription factor found in many transcription complexes (Althoefer et al., 1995; Elble and Tye, 1992; Jarvis et al., 1989; Kuo et al., 1997; McInerny et al., 1997; Oehlen et al., 1996; Passmore et al., 1988) that gains specificity through interactions with different sets of cofactors. ArgRII (Arg81) was identified as the arginine sensor (Amar et al., 2000) and ArgRI (Arg80) shown to confer specificity (Jamai et al., 2002), with Mcm1 providing affinity and ArgRIII (Arg82, IPMK) responsible for stabilizing the complex (El Bakkoury et al., 2000).

Simultaneous to investigations along the transcription complex line IPMK was independently identified as an I(1,4,5)P$_3$ and I(1,4,5,6)P$_4$ kinase by two groups while investigating inositol signaling (Saiardi et al., 1999; York et al., 1999). With this discovery a new realm of investigation opened into metabolism and inositol signaling. Not only was it now known that one of the components of the ARGR-Mcm1 complex was also a kinase, but that the small molecule it phosphorylated was already known to be involved in Ca$^{2+}$ signaling (IP$_3$) and when further phosphorylated played a role in mRNA
export from the nucleus. It also raised the question of how IPMK/Arg82 stabilized the ArgR-Mcm1 complex. Was it the result of IP production or a direct protein-protein interaction, or a combination of the two? From the kinase side a new set of questions arose – IPMK appeared to have a single active site, yet performed at least two different phosphorylation reactions in vivo. IP₃ was phosphorylated to I(1,4,5,6)P₄ and then to I(1,3,4,5,6)P₅ in an ordered 6/3-kinase pair of reactions. Just how many other IP substrates did it have and what were the roles of these various inositol polyphosphates?

Once identified as a multi-kinase, it was not long before homologues in other eukaryotes were reported (Chang et al., 2002; Fujii and York, 2005; Nalaskowski et al., 2002; Seeds et al., 2004; Stevenson-Paulik et al., 2002). With each discovery and the availability of synthetic IP species for testing, additional activities and IP substrates were identified for IPMK, with some differences in substrate profiles between homologues. The plant (A. thaliana) and yeast (S. cerevisiae) homologues thus far have the greatest range of IP substrates, yet even the more selective human homolog has 5-/3-kinase activity and weak 6-kinase activity.

Loss of IPMK is lethal in flies (unpublished data) and mice (Frederick et al., 2005). So far no eukaryotes other than yeast have been reported to survive without IPMK. As such most exploration of IPMK roles has been performed in yeast, where it is directly involved in at least arginine/ornithine metabolism (Messenguy, 1976; Messenguy et al., 1991) and several stress responses to salt and temperature (Caddick et al., 2007; Dubois et al., 2002; Zhu et al., 2009). Multiple other systems including mRNA export (Saiardi et al., 2000; York et al., 1999) are affected by IP species downstream of IPMK.
products. In yeast ipmkΔ cells the ability to utilize arginine or ornithine as the sole source of nitrogen is compromised, and cells are non-viable at 37°C. These phenotypes can be rescued with native or non-native IPMK, such as that from plant, human, and fly, leading to a perplexing series of questions. While the roles of different IP species can be partly elucidated by experiments using different IPMK homologues, the question remains in every case – which IP species is responsible for which phenotype?

**Questions**

What is the basis of IPMK selectivity? That is the principle question which inspired this work. The question of how an enzyme is able to distinguish between sixty-three different IP species (plain inositol is the sixty-fourth possibility) and not only recognize which were substrates, but also where to add phosphates to each, was most intriguing. Adding to the level of interest is the fact that IPMK accomplishes all of these reactions with a single active site, rather than having a separate site for each substrate and target hydroxyl (the 6’-, 3’-, or 5’-positions). While kinetic studies could identify which IP species were preferred substrates for which IPMK homolog (Chang and Majerus, 2006; Seeds et al., 2004; Stevenson-Paulik et al., 2002), they gave no mechanistic information regarding substrate selectivity.

How are phenotypes and IP species linked? Due to the multi-kinase abilities of IPMK it has been difficult to identify the relationships of specific IP species and phenotypes. If IPMK is removed entirely then an entire pool of various IP species is affected, either through accumulation or loss. While inositol-1,4,5-trisphosphate 3-
Kinase (IP3K) can be used to generate I(1,3,4,5)P₄ and therefore mimic the I(1,4,5)P₃ 3-kinase activity reported for the mammalian homologues (Fujii and York, 2005; Nalaskowski et al., 2002; Seeds et al., 2005), there are no reports of a kinase catalyzing only the 6-kinase reaction upon I(1,4,5)P₃. Therefore studies of IPMK have always been hampered by the uncertainty of if the results are the product of I(1,4,5,6)P₄ or I(1,3,4,5,6)P₅ generation, or the production or depletion of yet another IP species such as I(1,3,4,6)P₄. Previous investigations have suggested roles for the IP₄ versus IP₅ species, but evidence has remained inconclusive. Another debate has been ongoing – is the ArgR-Mcm1 complex and therefore arginine metabolism regulated through the poly-aspartate domain of ScIpmk or through inositol polyphosphate production? Countering theories have been proposed (Dubois et al., 2000; El Alami et al., 2003; Stevenson-Paulik et al., 2002; Xia et al., 2003), with evidence mounting that IP production plays the key role. It has also remained an open question which IP product of IPMK is responsible for temperature sensitivity.

**Objectives of this work**

_Determine the Structure of IPMK and Basis of Substrate Selectivity_ – To either directly answer these questions or develop the necessary tools, the first objective of this work was to determine the structure of IPMK and from it determine the basis of substrate selectivity. If the structure of a co-crystal containing an IP substrate (IP₃ or an IP₄ species) and ATP (or an analog) could be solved, then the basis of substrate selectivity might be identified. From this structural data it would be possible to move to the next stage of investigation.
Redesign IPMK to change its substrate profile – Knowing the basis of IP selectivity should make it possible to alter the binding site of IPMK to bind a different profile of IP species, or even turn it into a extremely specific kinase, such as IP3K which is specific for a single IP species and reaction. By generating multiple constructs with different IP profiles it should be possible to investigate inositol signaling and determine which IP species play which roles in vivo.

Investigate in-vivo inositol signaling – What roles do which IP species play in arginine/ornithine metabolism and temperature sensitivity? IPMK constructs with altered selectivity could be used to probe signaling in multiple organisms, but for the purpose of this work the goal was to focus on two of the most easily tractable phenotypes in yeast – utilization of arginine or ornithine as the sole nitrogen source or survival at 37°C. This would allow more rapid experimentation and refinement than investigations using more complex organisms such as flies or mice. Once the utility of the constructs was confirmed, they could then be used in other organisms.

Approaching Inositol Signaling with an Integrated Methodology

The underlying problem presented here is initially of a structural nature – how does IPMK distinguish between multiple inositol polyphosphate species? Each IP3, IP4, and IP5 species is identical in composition and electrostatic potential. The only differences are in the relative positions of the phosphate groups attached to the inositol ring – hence, the only differences are in their structures and stereochemistry. Furthermore the intent after determining the mechanism of IP selectivity was to alter this
mechanism by altering the structure of IPMK. Therefore the optimal approach for the first steps was a structural one. Protein crystallography was selected for two reasons. First, due to the size of the IPMK family members (32kDa or greater) NMR was likely to be more difficult than crystallography. Second, previous crystallization work had already been performed within the York lab and provided a starting point for further studies.

Once a structure had been solved and selective constructs designed, these would be tested through kinetic and *in-vivo* studies. While the structure of IPMK would be explain how the enzyme functioned, it would contain no information about catalytic efficiency or relative affinities, nor the *in vivo* roles of the different product inositol polyphosphate species. At this point it would become necessary to switch to an entirely different set of techniques including yeast genetics, metabolomics, and various phenotypic assays. Hence this work is an integration of multiple scientific disciplines ranging from the scale of the molecular (protein crystallography) through the protein complex (kinetics) to the entire organism (phenotypes).
Chapter 1. From Gene to Density; Solving the Structure of *Arabidopsis thaliana* Inositol Polyphosphate Multikinase α

*Introduction*

To collect protein crystallographic data a number of conditions must be met, the first of course being the generation of a protein crystal. This in turn has its own set of requirements. First, there must be a large amount of protein, often on the milligram or greater scale, of high purity. Any contaminating proteins or other materials can easily prevent crystallization by preventing the formation of interaction surfaces. Second, the protein must pack and crystallize in an ordered and regular fashion for the crystal to actually grow, and grow to a scale that is physically manageable, generally no smaller than the tens of micrometer level. Once a crystal has been grown it must be prepared for data collection, either at room-temperature or frozen, and then exposed to an x-ray beam and diffraction data collected. This chapter describes the various processes used to obtain diffraction quality crystals including redesign of the protein construct and a modified affinity purification protocol that enhanced purity while minimizing loss.

*First Constructs and Crystals*

Previous work had been performed in the York lab to crystallize the fly form of IPMK, but met with limited success. Crystals were difficult to reproduce and never grew to a manageable size. Two general approaches are often used in finding a construct that
will successfully crystallize – screening multiple constructs of a single species, or screening multiple species. Of course the approaches can be combined as well, but often crystallographic studies will begin with one of the two. The second approach was picked on the basis that with no prior IPMK structure solved and poor sequence identity to other IPMK family members (Stevenson-Paulik et al., 2002), there was little information to intelligently guide construct design. Four different homologues of IPMK were screened for protein production, using the criteria of total protein expression and percentage in the soluble fraction as the measure of success. The goal was to identify which species over-expressed the greatest in *E. coli* and had the greatest amount of soluble protein. These criteria were based on the understanding that any poorly expressing construct would not be expected to crystallize well, as a lack of stability in the cellular environment would likely be exacerbated in the crystallization environment. IPMK homologues from *Arabidopsis thaliana* (plant), *Drosophila melanogaster* (fly), *Rattus norvegicus* (rat), and *Saccharomyces cerevisiae* (yeast) were all expressed in *E. coli* BL21 DE3 (Stratagene) cells and screened as described in the experimental methods section. Each had already been cloned by previous works into the pGEX-KG vector, a modified pGEX-2T using an extended linker between the GST-affinity tag and MCS.

pGEX-KG uses the T7 induction system, a bacteriophage tRNA that allows for a high degree of control of expression using isopropyl-β-D-thiogalactopyranoside (IPTG) as the inducing agent. Briefly, a strain of *E. coli* that contains the T7 polymerase behind the Lac-UV5 promoter is transformed with the plasmid, which has the gene of interest behind a strong T7 recognized promoter. In the absence of lactose or its analog IPTG the
repressor lacI binds to the promoter region and prevents expression of the T7 polymerase. Upon addition of either, typically IPTG as it will not be metabolized, the repressor dissociates from the promoter and the T7 polymerase is expressed, which then moves to begin transcription of the gene of interest (Studier and Moffatt, 1986). Over-expression was measured by comparing the protein bands of the cellular lysates of induced and control cultures, with a large band expected in the former. A rough measure of stability could be determined by comparing the size of the induced band in the whole-cell lysate to the soluble fraction, separated by centrifugation. Loss in the intensity of the induced band relative to the other proteins present would indicate that some fraction of the protein was in the insoluble fraction. Further confirmation could be gained by resolublizing the cell pellet and running it on a gel compared to the uninduced pellet. Recombinant protein may then be purified by affinity chromatography using a glutathione sepharose column.

Additional parameters during expression also had to be taken into account. The temperature, length of induction, and concentration of IPTG may all influence the level of expression and degree of stability. While *E. coli* grow optimally at 37°C, that is not the optimal temperature for the stability of many proteins and therefore it was decided to induce the protein at 30°C after the cells had reached an optical density of ~0.6 at 600nm, as recommended by the product manuals. Expression over the course of induction was also measured, taking samples from each culture at regular time points to track protein production and determine if expression was lost or protein began to go to inclusion bodies after some time period. Finally, various concentration of IPTG were used, as each
construct may and likely will express differently and the optimal concentration of IPTG cannot be predicted.

As shown in Fig. 1-1, all GST-IPMK constructs were successfully over expressed in *E. coli* BL21 DE3 cells. It was also clear that over expression was far better for the plant and yeast homologues. Possible reasons include inherent stability differences, conditions of induction (particularly temperature in regards to the rat and human homologues), and the possibility of post-translation modifications that cannot be performed in a bacterial expression system. The latter was a consideration during the decision to proceed with expression in a bacterial system, as being a eukaryotic-specific system none of the inositol kinases exist in a prokaryotic organism. Protein screening was not performed in yeast as all of the constructs had previously been expressed and activity confirmed in bacterial systems (Frederick et al., 2005; Fujii and York, 2005; Seeds et al., 2004; Stevenson-Paulik et al., 2002). The question now was if yield, purity, and stability would be sufficient to the needs of protein crystallography experimentation.
Figure 1-1: IPMK Expression Screening

SDS-PAGE gels (12.5%) of the soluble fractions of *E. coli* BL21 DE3 transformed with pGEX-KG containing the plant, fly, rat, or yeast IPMK constructs. Lanes from left-to-right are ladder, uninduced, 1-5hrs post-induction time-points, and overnight.

With two clearly better constructs the decision was made to pick the plant homolog over the yeast for two reasons. First, the plant homolog, *At* Pmka or *at* IPK2α, appeared to have slightly higher levels of expression and soluble protein. Secondly, and more significant to long-term goals, the plant homolog appears to rescue yeast ipmkΔ phenotypes solely through its kinase properties (Stevenson-Paulik et al., 2002), while the yeast homolog is reported to have both protein-complex and kinase roles (El Alami et al., 2003). If inositol signaling in yeast were to be investigated using a redesigned IPMK,
then the plant homolog is preferable as effects would be due solely to the altered kinase properties. If the yeast homolog was used the question would remain as to whether or not the results were due to alterations in the kinase functionality or complex formation or stabilization, or some combination of these factors.

Using the optimized conditions identified by the construct screening GST-Atlpmkα (hereafter Atlpmk) was mass produced for use in crystallization trials. This consisted of growing large (6L) cultures of bacteria and purifying the protein as described in the experimental methods section. During lysis it was observed that a large amount of protein was in the insoluble fraction, though much remained soluble, and a significant fraction of protein was not binding to the affinity resin regardless of the amount of resin used. Most ominously, after cleavage of the GST affinity tag with thrombin the maximum concentration of Atlpmk dropped to only a few mg/ml, down from over 10mg/ml immediately prior to digestion. This was an immediate and serious concern. Protein crystallization often requires protein concentrations well in excess of 10mg/ml to grow any crystals, and by their very nature protein crystals will have a local concentration far higher than the solution they grow out of. A loss of solubility upon removal of extraneous material (the affinity tag) indicated a problem with the construct design. Not surprisingly this manifested itself in the failure of the cleaved protein to consistently grow crystals.

Several methods exist to grow protein crystals. At the non-robotic scale the two most common are vapor diffusion and batch or oil-immersion (robotic setups are almost exclusively micro-batch due to the technical details involved). Vapor-diffusion crystal
growth uses a two-component system, wherein a drop containing both protein and precipitant solution are suspended from a cover slip or sit in a small basin within a larger sealed well that contains a much greater volume of additional precipitant solution. Water moves from the smaller drop to the larger reservoir (vapor diffusion) to equalize the vapor pressure of the drop and reservoir. Movement is always to the reservoir due to the higher concentration of salts and other compounds in and greater volume of the precipitant solution than the crystallization drop. The rate of equilibration can be altered by temperature, drop volume and surface area as well as by the composition of the precipitant solution. As the volume of the protein-precipitant drop shrinks and the concentration of protein increases, crystal nuclei form that provide surfaces for additional crystal growth. Ideally the change in protein concentration is sufficiently slow that only a few crystal nuclei form and further movement of the protein from the soluble to solid stage occurs at a pre-existing crystal, rather than the formation of multiple small nuclei. Batch or oil-immersion setups simply place the protein and precipitant together under a layer of oil to prevent evaporation. There is no gradual change in the concentrations of the various components in solution, decreasing the conditions sampled in each experiment. The advantages of the batch setup include reduced amounts of protein and precipitant solution required and often faster setup, as well as being more amenable to automation.
Hanging-drop vapor-diffusion setups were performed with cleaved \textit{At}Ipmk using the Nextel Classic Screens I and II precipitant solution sets. Crystals were observed in a single condition; 0.1M HEPES pH7.5, 2\% (v/v) PEG400, and 2M (NH$_4$)$_2$SO$_4$ (Fig. 1-2).

\textbf{Figure 1-2: Initial \textit{At}Ipmk crystals}

Attempts to improve crystal growth through the use of additives, overlaying the drops with mineral oil (to slow the rate of equilibration), and varying the protein and precipitant concentrations and drop size slightly improved crystal size, from 80x80x40\,µm stacked plates to a maximum of 140x120x120\,µm as single crystals. Unfortunately the resolution never exceeded 3.3\,Å, insufficient to solve the structure and the crystals proved to be very sensitive to radiation damage during data collection. Complete datasets were collected for a few crystals, at which point a new challenge emerged.

\textit{Encountering Twinning}

While complete datasets were eventually collected out to 3.3\,Å, molecular replacement using the core of \textit{Hs}Ip3k (PDB accession code 1W2F) as a search model failed. It was initially assumed that the inability to find a phase solution was due simply
to the still poor resolution of the data, until anomalous diffraction data was also analyzed. On the likelihood that molecular replacement would fail, given the lack of homologous proteins with structures, seleno-methionine labeled *At*1pmk was also produced and crystallized in similar conditions to the native construct. Anomalous diffraction is named for the increased anomalous signal strength of larger atoms (sulfur or larger) that may be used to solve their positions and therefore extended to solve the phases of the entire dataset. By collecting data from protein expressed in conditions that resulted in seleno-methionine substitution for regular methionine, each protein would carry within it heavy-atoms that could generate sufficient anomalous diffraction differences to be detected and used to solve the structure.

However, despite anomalous signal to 3.4Å a phase solution using SHELX (Sheldrick, 2008) could not be calculated. Analysis of the reflection data with the Merohedral Crystal Twinning Server, UCLA (http://nihserver.mbi.ucla.edu/Twinning/) revealed that half the native crystals and the sole successfully collected seleno-methionine crystal were twinned. Twinning is the physical phenomenon in crystals where at least two separate but symmetrically related lattices exist in the same crystal (Fig. 1-3).
Figure 1-3: A twinned crystal with two lattices

As both lattices will generate diffraction data and are related to each other, deconvoluting the reflections can be impossible. Unless the degree of twinning, the twinning fraction, is either very small or very close to 50%, accurately assigning each reflection the correct assignment and percent contribution from the overlapping lattice points becomes very difficult. At low levels one lattice can simply be ignored and treated as noise, while at 50%, known as perfect merohedral twinning, the contribution of each lattice to each reflection is always 50% (Yeates, 1997). Unfortunately all the twinned crystals had twinning fractions in the 30-40% range. Efforts turned to attempts to grow improved crystals rather than attempting to detwinn still poorly diffracting (~3.5Å) crystals.

**Switching to a new tag, linker, and construct**

Combined with the issues of protein expression and solubility already observed, clearly changes to the purification process and a new construct were needed. While
twinning is a phenomenon of crystal packing and may be overcome by crystallization in
different conditions, packing is also ultimately influenced by the structure of the protein
itself, dictated by the residue sequence. Therefore both the production of *Ar*Ipmk and the
constructs themselves were redesigned.

Previous purifications of *Ar*Ipmk used relatively low levels (50mM) of potassium
chloride in the lysis, digest, and purification buffers. In protein crystallography it is
common practice to keep the concentrations of any buffer components as low as possible
to minimize interaction with precipitant conditions and maximize the probability of
crystal growth. A high concentration of a solubilizing compound, while ideal for high-
concentration solutions, may also inhibit crystal growth when the goal is to drive the
protein *out* of solution. However, if the concentrations of buffering components are too
low the protein may simply be insoluble. The single step of increasing the concentration
of potassium chloride to 500mM during lysis and 150mM during affinity column
purification and concentration increased solubility to ~10mg/ml and in the same
crystallization conditions resulted in crystals that grew to 600x400x300µm.

Simultaneously a new affinity purification scheme was devised to both reduce the
amount of time required to purify affinity-tagged proteins and reduce the level of
background contamination. During elution of the tagged-enzyme, even when using a
gradient a number of native *E. coli* proteins co-eluted with IPMK, including some of
similar molecular weights. While it might be possible to remove the proteins
significantly larger or smaller than IPMK with size-exclusion gel-chromatography,
proteins of similar size would be expected to remain. It was also noted that the longer the
purification process took the more protein was lost at each step, presumably due to denaturation over time. Digesting the resin-bound protein directly and then washing with buffer lacking any eluting compound (glutathione for a GST-tag) failed to significantly reduce the amount of contamination. Though not a large amount, any contamination could easily interfere with crystallization. To address this concern, a new protocol utilizing thrombin digestion and a serial series of affinity columns was developed. After loading and washing the resin the column was resuspended in one volume of thrombin digestion buffer and transferred to an appropriately sized syringe sealed with a stop-cock. Thrombin was added and excess air removed, then the stop-cock closed and the syringe set on a rotator overnight at 4°C. This eliminated the time required to first elute and then concentrate the protein prior to thrombin treatment and the need for dialysis to remove the glutathione.

The following day the resin was transferred to a new column and connected in-serial to a fresh glutathione sepharose plus p-amino benzenidine sepharose column. This system was designed to remove in one pass both the thrombin and any residual native proteins with low glutathione affinity. During the overnight digestion $Atl\text{pmk}$ would be freed from the column, leaving the tag attached, and the now solution suspended protein could be washed from the column. However, continuing residual $E.\ coli$ contamination required an additional step. Adding a second glutathione column provided an environment that would be able to bind any low-affinity residual proteins, and without competition from the now cleaved and glutathione-affinity lacking $Atl\text{pmk}$. Thrombin would be removed by the p-amino benzenidine sepharose resin, since at ~35kDa it could
not be expected to be reliably separated from the 32kDa \textit{At}Ipmk by gel chromatography. The suspended protein, from the column volume of buffer added prior to digestion, was collected in fractions from the dual-column setup and a second set of fractions collected running additional buffer over the system. Each fraction was assayed for the presence of protein and gels run of the protein-containing fractions. Surprisingly, not only was the purity of the protein significantly increased, but it was increased to the point that even when concentrated and very large amounts were loaded no band other than that corresponding to the size of \textit{At}Ipmk was observed. Protein purified over the double-affinity column system was more pure than that purified over one affinity column and one gel-chromatography column, and with less loss of protein due to dilution effects and aging. To further confirm purity the pooled and concentrated protein was run over an S200 gel chromatography column and elution over time measured by UV spectroscopy. A single peak was observed at approximately 65kDa, corresponding to a dimer of 32.2kDa monomers. The purified protein could be concentrated the maximum 10mg/ml previously stated when 150mM potassium chloride was present.

Sadly crystals grown in the high-salt buffer also proved to be twinned. This clearly indicated a packing problem that was suspected to be a characteristic of the construct itself. Although previous crystals had been twinned roughly half the time, none of the untwined crystals had provided sufficient quality data to solve the structure of \textit{At}Ipmk. While contamination had been reduced and eliminated as a suspect in the twinning problem and far larger crystals could be grown, hopefully to provide higher
resolution diffraction though the simple matter of more atoms in the x-ray beam, a solution had still not been found. This indicated a need to revisit the original construct.

Examination of the pGEX-KG vector indicated several potential sources of the issues with protein solubility and failure to grow useful crystals. As can be seen in the pGEX-KG vector map (Fig. 1-4) the GST tag is attached to a long, nineteen-residue linker that contains eight glycine residues, five in a row, and several additional hydrophobic residues. The center of the linker is highly hydrophobic with the sequence PGISGGGGGII.

FIGURE 1-4: The pGEX-KG cloning site

The thrombin cleave site (arrow), linker, and cloning sites used to insert AtIpmk. Note the length and composition of the residues between the thrombin cleavage and Ncol insertion sites.

Given that GST and IPMK are soluble proteins, it is obvious that a long, hydrophobic linker may introduce a solubility defect. Secondly, and of more concern, the thrombin cleavage site is placed on the opposite side of the linker from the MCS. Therefore upon cleavage of the GST tag there would remain at the end of IPMK a large, fifteen-residue hydrophobic region now lacking even the effect of GST to keep the entire construct soluble. The option of attempting to crystallize the entire fusion construct was declined for two reasons. One, any additional residues, especially an entire protein such as GST, can introduce artifacts into the structure of a protein. The only way to be sure of
avoiding such artifacts is the removal of any artificial elements. Two, crystallization requires a regular, repeating unit of packed protein – a long, flexible linker between two globular proteins could easily interfere with packing by introducing too much relative positional flexibility between the two domains. It might also play a role in the observed twinning by providing an end that could adopt multiple conformations that would affect packing.

Clearly at least two changes were required. First, a shorter, non-hydrophobic linker was needed to replace the current one, and second, the thrombin cleavage site needed to be moved to between the linker and IPMK. Fortunately two such vectors were immediately available. The Gewirth lab at Duke, Biochemistry department, had the pGEX-Nb and pET-15b vectors and generously supplied aliquots for use. pGEX-Nb is another derivative of pGEX-2T, containing the GST affinity tag, a short linker, thrombin cleavage site, and a much smaller multiple cloning site that contains only the NdeI-BamHI sites. pET-15b uses a 6xHistidine tag instead of a GST tag. Both vectors have identical cloning sites and are designed such that upon cleavage with thrombin only a Gly-Ser-His remainder (part of the thrombin cleavage and NdeI sites) is left on the N-terminus of the target protein (Fig. 1-5).

![Thrombin sequence](image)

**Figure 1-5: pET-15b and pGEX-NB cloning site**
The cloning sites of pET-15b and pGEX-NB are identical. Contrast the proximity of the thrombin cleavage site to that of the pGEX-KG vector in Figure 1-4. After cleavage only a GSH should remain.

*Ipmk was cloned into each vector as described in the experimental methods section and both constructs screened for expression and solubility before and after treatment with thrombin. The other species of IPMK were not screened a second time on the assumption that as the only changes being made were in the linker and affinity tag, the relative stabilities of the first set of pGEX-KG constructs would hold true were they all to be cloned into the pGEX-Nb and pET-15b vectors.

As predicted both constructs had significantly improved expression and solubility. Nearly no protein was observed in inclusion bodies, essentially all soluble protein was bound the affinity columns, and there was no loss of solubility after cleavage of the affinity tags. Both constructs, identical after cleavage, could be concentrated to over 20mg/ml in 50mM Tris-HCl pH7.5, 150mM KCl, and 5mM β-Me. Expression levels of the 6xHis-tagged construct were slightly higher than the GST-fusion construct, likely due to the greater amount of time and resources required for the cells to generate the twice-as-large fusion construct. All future protein production efforts therefore shifted to using the pET-15b construct (Fig. 1-6).
The end goal of screening and optimizing the protein generation process was to grow crystals suitable for crystallographic data collection. Once again the Nextel crystal screens were used to search for conditions that *Atlpkm* would crystallize in. Of the ninety-six different conditions tested, one produced crystals. Crystals were observed in 0.1M Tris-HCl pH 8.0, 2.0M (NH₄)₂SO₄ and 10mg/ml *Atlpkm* at 17°C. Further expansion of drop size, ratio of protein-to-precipitant volume, protein concentration, and pH and salt concentration led to a final crystallization condition range of 0.1M Tris-HCl pH 8.5, 2.0-2.15M (NH₄)₂SO₄, and 16-20mg/ml *Atlpkm* at 17°C. Crystals grown in these conditions commonly reached 600x400x300µm, more than sufficient size for mounting and data collection. Seleno-methionine labeled full-length *Atlpkm* was also generated using the pET-15b construct and incorporation confirmed (Fig. 1-7).

Unfortunately data collected from these short-linker, full-length *Atlpkm* crystals proved to be twinned as well. In anticipation of this possibility truncation constructs at the N-terminus were also designed during cloning into the pGEX-Nb and pET-15b
vectors. Shortly after discovering the twinning issue the structure of the yeast IPMK homolog was published at 2.0Å (Holmes and Jogl, 2006). Observing that the N- but not C-terminus was disordered new constructs were designed that deleted seven, sixteen, or twenty-five residues from the beginning of AtIpmk. These positions were picked as having hydrophilic residues leading groups of hydrophobic residues, on the understanding that leaving a hydrophobic sequence exposed was likely to impair solubility and crystallization. Protein was successfully produced for the sixteen and twenty-five residue truncations (hereafter Δ16 and Δ25) and again screened. Crystals of the Δ16 construct grew in the same conditions as full-length protein (Fig. 1-8) and complete datasets were collected to 2.9Å at the Advanced Photon Source at Argonne Chicago. While some crystals were twinned, most were not.
Figure 1-7: MALDI-ToF Spectrum of Sel-Met and Native Atlpmk

Figure 1-8: Final 16-286 Atlpmk crystals
Solving the Phase Problem

As previously mentioned a phase solution was not found using the cores of $HsI$ or $ScI$ as the search models for molecular replacement. Efforts turned to *de novo* phasing use seleno-methionine substituted $AtI$ to determine experimental phases and solve the structure. This work used the Single-Wavelength Anomalous Diffraction technique to solve the structure of $AtI$.

As its name implies, SAD phasing uses a single wavelength to collect the necessary data, and hence a single data set is collected. Multiple datasets collected at the same wavelength will increase redundancy, but not provide additional *unique* data. Datasets collected at multiple wavelengths may be used to calculate a Multi-wavelength Anomalous Diffraction (MAD) phase solution. During data collection all $AtI$ crystals were found to be highly radiation sensitive, visibly becoming discolored over the course of data collection. Sensitivity may have been due to the high solvent content which would have provided more water molecules for radical formation that could damage the crystal and less protein so that any damage was of greater relative effect. Seleno-methionine labeled crystals were particularly sensitive and it was not possible to collect multiple datasets from a single crystal at different wavelengths to calculate a MAD phase solution. Attempts to collect different wavelengths from different crystals did not succeed as radiation damage rapidly accumulated and only one dataset was useable, a single full-length crystal diffracting to 3.1Å. With only a single useable anomalous diffraction dataset the structure of $AtI$ could only be solved by SAD phasing. A disadvantage of SAD phasing is that two possible phase solutions exist, as
enantiomers of each other, and cannot be distinguished by examination of the diffraction data. Identification of the correct “hand” solution requires examination of density maps for interpretable density. MAD experiments avoid this problem by collecting data at several wavelengths which eliminates the ambiguity, but of course the collection of multiple datasets may present additional challenges, as seen in the case of AliPmk.

Analyzing the anomalous diffraction data with the SHELX program suite (Sheldrick, 1997; Sheldrick, 2008) identified four possible solutions. In addition to the phase enantiomer ambiguity, the reflection data could be indexed into two space-groups; P3\textsubscript{2}121 and P3\textsubscript{1}21. Examination of the diffraction data alone cannot distinguish between these two space-groups as the 3\textsubscript{2} and 3\textsubscript{1} screw-axis have the same reflection systematic absences. As in determining the correct phase solution, the only way to distinguish between these space groups is to calculate and compare the electron density maps to find the solution that generates continuous electron density that follows known rules of protein structure, such as residue chirality. Therefore it was necessary to examine four different experimental density maps; each spacegroup with each phase solution. Comparison of these maps clearly revealed the correct phase solution and identified the spacegroup as P3\textsubscript{2}121 (Fig. 1-9). The maps for the other three solutions lacked any continuous density or clear delineations of protein versus solvent regions and were distorted.
Building and Refinement

Once the structure of Atlpmk had been solved the next step was to build and refine a model. This proved to be a challenging task due to the limited resolution of the data. At only 2.9Å the level of detail visible in the initial electron density maps was limited, with density for many side-chains difficult to interpret or simply not observed. Similarly a number of loops had no observable electron density, indicating these regions were disordered and therefore could not be modeled.
As expected the core density seen in the experimental maps was very similar to
the core of the structures of ScIpmk and HsIp3k. This provided an avenue to quickly
build part of the AtIpmk model. The conserved region of the ScIpmk model was
converted to a poly-alanine model and fit into the experimental density map using the
Molrep program from the CCP4 suite (1994). Simulated annealing using the
Crystallographic and NMR System (CNS) program (Brunger et al., 1998) was used to
initially refine the poly-alanine model coordinates. Using the seleno-methionine
anomalous difference density maps, where density would be observed only for the
selenium atoms replacing the sulfur of the methionine residues, the register could be
assigned from the primary amino-acid sequence. Parts of the electron density were
discontinuous from those that contained the selenium peaks and were assigned based on
the characteristic density patterns of larger residues such as histidine, phenylalanine,
tryptophan, and tyrosine.

Additional residues were built in the O program (Jones et al., 1991) to the limit of
the observable density and a second round of simulated annealing performed to refine the
coordinates. Throughout the model building and refining process a combination of
experimental, simulated annealing composite-omit, 2F_oF_c, and F_oF_c maps were used.
After a model had been built for all the experimental density a final round of simulated
annealing was performed and a new composite-omit map calculated. Further model
rebuilding continued in O, but now also used the KiNG and MolProbity programs (Chen
et al., 2009; Davis et al., 2006; Davis et al., 2004; Lovell et al., 2003) to further refine the
rotamer and Ramachandran parameters. Refinement was then performed using the
positional minimization tools in CNS. Later translation-libration-screw (TLS) refinement (Winn et al., 2001) and the REFMAC5 program (1994) from the CCP4 suite were used to finish refinement of the \( A1p \) model, as determined by convergence to its final \( R/R_{\text{free}} \) values of 23.64/24.61 and modeling of all interpretable density.

**Summary**

Solving the crystal structure of \( A1p \) required a long process that involved several construct redesign steps and multiple alternative techniques to overcome difficulties in generating protein and crystals. Several different homologues were screened for sufficient expression levels only to find that the original expression vector was unsuitable to the needs of crystallography. After transfer into a more suitable vector large crystals could be grown, but contaminating proteins continued to prove difficult to eliminate until a new multi-affinity column protocol was developed. Extremely large crystals could be grown from the now highly purified and soluble protein, but were found to be twinned, a crystal packing phenomenon that rendered them useless. Truncating the N-terminus, observed to be disordered in the yeast homolog, resulted in the 16-286 (\( \Delta 16 \)) construct that over-expressed and purified to a greater level than any previous construct and grew similarly sized crystals. Datasets collected of the \( \Delta 16 \) construct diffracted to 2.9Å, but a phase solution could not be found by molecular replacement using the core structures of either \( HsIp3k \) or \( ScIp \). *De novo* phase information was collected by single-wavelength anomalous diffraction analysis of seleno-methionine crystals. Due to the sensitivity of the \( A1p \) crystals to radiation damage a single anomalous scattering
dataset was successfully collected and density maps for four possible solutions were calculated. Examination identified the correct enantiomer and spacegroup (P3_21) and part of the core of the structure of ScIpmk was modeled into the experimental density as a starting point. The rest of the model was completed by building in O and refining with the programs CNS and REFMAC and quality analyses were performed using MOLPROBITY and KiNG. An examination of the products of these labors are presented in the following chapter.

**Experimental Methods**

*Induction screening – E. coli* BL21 DE3 cells were transformed with the pGEX-KG vector containing either AtIpmk, DmIpmk, RnIpmk, or ScIpmk and plated on LB-ampicillin media. Individual colonies were used to inoculate 25ml liquid cultures and grown to an absorbance of 0.6 at 600nm, 37°C and expression induced with 0.01mM, 0.1mM, 0.2mM, 1.0mM, and 10.0mM final concentration isopropyl-β-D-thiogalactopyranoside (IPTG). The temperature was reduced to 30°C and 5ml fractions were collected at 1-5 hours post-induction as well as immediately before induction and after overnight growth. Each sample was pelleted and resuspended in 50mM Tris-HCl pH7.5, 50mM NaCl, and 5mM β-mercaptoethanol (β-Me) and lysed by sonication for 5x1sec cycles. The soluble fraction and cellular debris were separated by centrifugation at 16,100 x g for 30min and samples loaded onto Amersham 12.5% PhastGels. After separation each gel was stained with Coomassie R-250 and examined for induction of the target proteins over time and compared between species and concentration of IPTG.
Plasmid Constructs – Full length and N-terminal truncated (residues 16-286; Δ16)

AtIpmk were amplified from the pGEX-KG construct (Stevenson-Paulik et al., 2002) with restriction sites Ndel (N-terminus) and BamHI (c-terminus) and cloned into the 6xHis-tagged vector pET-15b and GST-tagged vector pGEX-Nb. The primers were: full-length primer sense, 5’-CAT AGT GGA CAT ATG CAG CTC AAA GTC CCT GAA CAT CAG-3’, 16-primer sense, 5’-CAG TAT TAT AAA GAC GGG AAG CCT GGT CCT CTC-3’ and antisense, 5’-TAC GGA TCC TCA TCA CTA AGA ATC TGC AGA CTC ATC TGC-3’. The constructs were transformed into XL1-blue chemically competent cells and plasmid extracted from transformants and sequenced at the Duke University DNA sequencing facility.

Expression and Purification of Final Recombinant 6xHis and GST-Fusion Proteins – E. coli BL21 DE3 (Invitrogen) cells were transformed with the pGEX-Nb, or pET-15b vectors containing AtIpmk. 25ml LB media plus 0.1mg/ml ampicillin was inoculated from a single colony of BL21 DE3 E. coli transformed with pET-15b AtIpmk or pGEX-Nb AtIpmk and grown overnight at 37°C, 250RPM. 4x1.5 liters LB media with 0.1mg/ml ampicillin were inoculated with 5ml overnight culture at 37°C and grown to an absorbance of 0.6 at 600nM. Expression was induced by the addition of 1mM IPTG (final concentration) for one hour and the temperature reduced to 30°C and grown overnight. Seleno-methionine labeling was performed by the methionine inhibition process (Van Duyne et al., 1993) and complete incorporation confirmed by mass spectrometry. M9 minimal media was prepared by combining 200ml 5xM9 salts ((64g Na₂HPO₄, 15g KH₂PO₄, 2.5g NaCl, 5g NH₄Cl)/L), 2ml of 1M MgSO₄, 20ml of 20%
filter sterilized glucose, 100μl of 1M CaCl₂, and bringing to 1L with sterile water. 50ml M9 media with 0.1mg/ml ampicillin was inoculated from a single colony from a fresh BL21 DE3 E. coli pET-15b Adpmk plate and incubated at 37°C approximately 18 hours. 10ml overnight culture was used to inoculate 4x1.5L M9 plus antibiotic media and grown at 37°C, 250RPM until an OD of 0.6 at 600nm had been reached. Each culture was then induced with 1.0mM final concentration IPTG and grown another three hours at 30°C, 250RPM. Cells were harvested by centrifugation in a Sorvall RC-5C Plus centrifuge at 3300 x g for 10min.

Cell pellets were resuspended in 50mM Tris-HCl pH7.5, 500mM KCl, 5mM imidazole, 5mM β-Me and 1 tablet "Complete Mini" protease inhibitor (Roche) per 10g of cell pellet and lysed in three passes at 15,000 PSI through a M110L microfluidizer (Microfluidics). Cellular debris were pelleted by centrifugation at 29,000 x g for 30 min and the supernatant loaded onto a Ni-NTA Superflow or glutathione-sepharose column (Qiagen) pre-equilibrated with 10 column volumes (cv) of lysate buffer. The column was washed with 5cv lysis buffer and 5cv digestion buffer (50mM Tris-HCl pH7.5, 150mM KCl, 5mM β-Me) and resuspended with one volume of digestion buffer. 17.8μg human thrombin (Haematologic Technologies) was added per ml affinity resin and digested overnight on a rotator. The resin was reloaded and a second column of the appropriate affinity resin over p-amino Benzamidine Sepharose 4 Fast Flow (GE Healthcare) resin attached in serial. The sample from the first column was run over the second and fractions collected and both columns washed with additional digestion buffer and a second fraction set collected. Fractions with protein were identified by spot-testing with
Coomassie R-250 dye (1-3μl spotted onto blotting paper and stained with Coomassie) and were pooled and concentrated using Microcon YM-10 10kDa micro-centrifugal concentrators (Amicon) and run over a Superdex 200 HiLoad 16/60 gel filtration column (Pharmacia). Final purity was confirmed by SDS-PAGE gels and pure protein pooled and concentrated to 10mg/ml or greater by absorbance at 280nm ($\varepsilon_{280} = 33015 \text{ M}^{-1} \text{ cm}^{-1}$ for wild-type, K100A, and 16-286, 35560 M$^{-1}$cm$^{-1}$ for K117W and K121W, 41250M$^{-1}$ cm$^{-1}$ for K117W:K121W), aliquoted and flash-frozen with liquid nitrogen, and stored at -80°C. All purification stages were performed on ice or at 4°C.

**Maldi-ToF Mass Spectroscopy** – Incorporation of seleno-methionine was confirmed by Matrix Assisted Laser Desorption/Ionization-Time of Flight mass spectroscopy using an Applied Biosystems Voyager-DE biospectrometry workstation. Samples were prepared at 10mg/ml of native and labeled protein and mixed 1:50 with a saturated 3,5-Dimethoxy-4-hydroxycinamic acid (sinapinic acid), 50% acetonitrile, and 0.1% tri-fluoroacetic acid solution. 1μl was spotted onto a stainless steel MALDI target and allowed to evaporate at room temperature. 30 nanosecond laser pulses were averaged to obtain a spectrum. A mass difference of 46.89Da per seleno-methionine was expected (Se: 78.96, S: 32.07), for a total of 281.34Da. The observed difference was 313.79, an excess of 32.45Da. The cause of the discrepancy is unknown; resequencing of the construct did not reveal any mutations and the same cell stocks were used to grow the native and labeled cultures.

**Crystallization and Data Collection** – Native Δ16 atIPK2α crystallized in 2.0-2.15M ammonium sulfate and 0.1M Tris-HCl pH8.5 at 16-20 mg/ml. Full-length seleno-
methionine atIPK2α crystallized in 2.10M ammonium sulfate and 0.1M Tris-HCl pH7.0 at 18 mg/ml. All crystals were grown by hanging-drop vapor diffusion using Hampton hanging-drop trays. Crystals were cryo-protected by successive transfer into 30% glycerol and saturated ammonium sulfate at the matching pH in 5% glycerol, 5min soak increments and equilibrated overnight in the final protectant, then flash-frozen with liquid nitrogen. Data were collected at the Advanced Photon Source, Argonne National Laboratory (Argonne, IL), beamlines 22ID (Seleno-Methionine) and 22BM (native) on MAR300 or MAR225 CCD detectors, respectively.
Chapter 2. The Structure of *Arabidopsis thaliana* Inositol Polyphosphate Multi-Kinase-α and a New Substrate Selectivity Theory

**The Structure of Arabidopsis thaliana IPMK**

Introduction

The structure of *Arabidopsis thaliana* Inositol Polyphosphate Multi-Kinase α was solved to 2.9Å with an R/R<sub>free</sub> of 23.64/24.61 and bond and angle RMSD values of 0.015Å and 1.496°. Each asymmetric unit contains two copies of AtIpmk, mirroring its elution as a dimer in gel chromatography (data not shown). The N-terminus and several loops are missing, with 215 of 286 residues are modeled in each monomer, consisting of residues 41-70, 81-175, 183-230, and 238-279. Anomalous diffraction density is observed for five of six seleno-methionine residues, the missing methionine is the first residue and in the disordered N-terminus. Only a few residues have poor backbone geometry, 92.5%/99.5% are in the favored/allowed regions of the Ramachandran plot. The solvent content is abnormally high at 76.5% with V<sub>m</sub>=5.22 (Matthews, 1968), but experimental phase maps reveal no density for additional molecules nor can any be modeled within the constraints of the asymmetric unit dimensions and symmetry function. Anomalous difference maps do not have any peaks outside those in the two monomers. Additional statistics are presented in Table 2-1.
# Table 2-1: Crystallographic Data, Phasing, and Refinement Statistics

## Data Collection and Phasing

<table>
<thead>
<tr>
<th></th>
<th>16-286</th>
<th>SeMet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space Group</td>
<td>P3₂1</td>
<td>P3₂1</td>
</tr>
<tr>
<td>Unit cell (Å)</td>
<td>a = b = 130.8</td>
<td>a = b = 130.2</td>
</tr>
<tr>
<td></td>
<td>c = 129.9</td>
<td>c = 129.4</td>
</tr>
<tr>
<td></td>
<td>α = β = 90°</td>
<td>α = β = 90°</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.000</td>
<td>0.97166</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>46-2.9</td>
<td>50-3.1</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>28738</td>
<td>44459</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>100 (100)²</td>
<td>100 (100)</td>
</tr>
<tr>
<td>R&lt;sub&gt;merge&lt;/sub&gt;</td>
<td>8.5 (52.4)</td>
<td>9.2 (52.6)</td>
</tr>
<tr>
<td>FOM (SAD)</td>
<td></td>
<td>1.31</td>
</tr>
</tbody>
</table>

## Refinement

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>46-2.9 (2.975-2.90)</td>
</tr>
<tr>
<td>No. of reflections</td>
<td>314752 (43621)</td>
</tr>
<tr>
<td>R (%)</td>
<td>23.64 (31.5)</td>
</tr>
<tr>
<td>R&lt;sub&gt;free&lt;/sub&gt; (%)</td>
<td>24.61 (37.0)</td>
</tr>
<tr>
<td>RMS deviations:</td>
<td></td>
</tr>
<tr>
<td>Bond length (Å)</td>
<td>0.015</td>
</tr>
<tr>
<td>Bond angle (°)</td>
<td>1.496</td>
</tr>
<tr>
<td>B factor (Å²) main chain bonds</td>
<td>0.917</td>
</tr>
<tr>
<td>B factor (Å²) side chain bonds</td>
<td>2.725</td>
</tr>
<tr>
<td>Protein atoms, number</td>
<td>3444</td>
</tr>
<tr>
<td>Sulfate atoms, number</td>
<td>30</td>
</tr>
<tr>
<td>Residues in allowed φ-ψ region</td>
<td>99.5%</td>
</tr>
</tbody>
</table>

---

² Data in parentheses are for highest resolution shells

<sup>b</sup> \( R_{merge} = \Sigma_{hkl} \sqrt{(N(N-1)) \Sigma_i |I_i(hkl) - <I(hkl)>|/\Sigma_{hkl} \Sigma_i I_i(hkl)} \)
Structural Description

As expected AtIpmk has a similar secondary and tertiary structure to other members of the ATP Grasp fold family (Fan et al., 1995; Murzin, 1996) consisting of eight α-helices and ten β-strands organized around a central β-sheet backed by slightly off-parallel α-helices and organized into three domains (Fig. 2-1). The N- and C-terminal sub-domains are α+β groupings while the inositol binding domain consists of two α-helices and connecting loop regions. The secondary structure assignment by Holmes and Jogl for the yeast structure is followed to facilitate comparisons. The N-terminal domain contains residues 41-85 (residues 1-40 are disordered) and is composed of two anti-parallel β-strands and two α-helices. The β-strands are designated as β2 and β3 as they align with the correspondingly numbered strands in the yeast structure. While density was observed in the volume predicted to be occupied by the first β-strand, it was of insufficient quality to model. Helices α1 and α2 precede the β2-strand and likely connect it to the presumably partially ordered β1-strand.

The larger C-terminal sub-domain makes up the majority of the structure of AtIpmk. Residues 94-102 and 132-277 are organized into five anti-parallel β-strands in a 3-2-1-4-5 order flanked by helix α6 and backed by helices α7, α8, and α10 and two smaller anti-parallel β-strands.
Figure 2-1: Structure of AtIpmk

A. Structure of AtIpmk. Stereo view with secondary structure elements named to match ScIpmk.
B. Representative density. Stereo view of Fo simple map density at 2\(\sigma\) and anomalous difference density at 8\(\sigma\).

Yeast helices \(\alpha5\) and \(\alpha9\) are not observed in the plant structure. Both are within inserts unique to the yeast sequence. Helix \(\alpha5\) is replaced by a short linking loop while \(\alpha9\) is within the insert replaced by a short but disordered region between residues 231-237. Similarly the \(\beta6\) strand is part of the same insert as helix \(\alpha5\) and not present. This insert is thus far unique to ScIpmk as it not present in HsIpk3k, which has the same structure in this region as AtIpmk. The inositol binding domain is inserted between
strands β4 and β5. Helices α6, α7, and α8 connect β6 and β7 while the small sheet strands β8 and β11 are between α8 and β9, and β10 and α10, respectively.

**Figure 2-2: Structure-based sequence alignment of Atpmk and ScIp mk and Hsp3k**

Uniquely helix α8 is split into two helices designated α8a and α8b by a three-residue insert (194-196) that loops and protrudes outward from the protein core (Fig. 2-3.A). Though the sequence of helix α8 is poorly conserved, both ends align very closely with the yeast and IP3K helices and are on the same axis, hence designation as parts of a single helix. The extended loop forms a sulfate salt-bridge and has direct hydrogen-bond contacts (Table 2-2) with the same chain symmetry mate (Fig. 2-3.B). The N- and C-domains are connected by the loop residues 85-93.
Figure 2-3: Helix α8 loop and interface

A. Packing at the extended loop of helix α8. Stereo image of helix α8 and bridging sulfate ion. Original structure in green, symmetry mate in orange.

B. Direct and salt-bridge interactions. Stereo image close-up of packing interactions at and surrounding the extended loop. All maps are 2F_oF_c at 1.5σ.
Table 2-2: Hydrogen Bonds at helix $\alpha_8$ interface

<table>
<thead>
<tr>
<th>Atom 1</th>
<th>Atom 2</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Salt Bridges</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SO$_4$ 501 O1</td>
<td>Ser195 N</td>
<td>2.99$^1$</td>
</tr>
<tr>
<td>SO$_4$ 501 O1</td>
<td>Gly194 N (symm)</td>
<td>2.81$^1$</td>
</tr>
<tr>
<td>SO$_4$ 501 O2</td>
<td>Ser189 O$_\gamma$ (symm)</td>
<td>3.97$^1$</td>
</tr>
<tr>
<td>SO$_4$ 501 O4</td>
<td>Ser195 O$_\gamma$</td>
<td>3.58$^1$</td>
</tr>
<tr>
<td><strong>Direct Contacts</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser195 O</td>
<td>Thr200 O$_\gamma$ (symm)</td>
<td>3.67$^1$</td>
</tr>
<tr>
<td>Thr200 O$_\gamma$</td>
<td>Thr200 O$_\gamma$ (symm)</td>
<td>4.09$^1$</td>
</tr>
</tbody>
</table>

$^1$ Mirror interaction across the 2-fold symmetry fold
symm – symmetry mate molecule

The inositol binding domain is designated by structural homology to the $Hs$Ip3k inositol binding domain and inserted between the C-domain $\beta$4 and $\beta$5 strands. As in the yeast structure the $\alpha$3 and $\alpha$4 helices match two of the $Hs$Ip3k helices, but lack the additional structural elements responsible for the greater selectivity of IP3K.

Packing between non-crystallographic copies is minimal, limited to parts of the N-terminal domain (residues 54-58, 68) with a few residues in the C-terminal domain (195-196, 203, 274, 277). Crystallographic packing is much more extensive, with contacts throughout the $\alpha$-helix packed side of the central $\beta$-sheet in a back-to-back arrangement where half of the helices are interact with one monomer and half with the other. A second set of contacts exist at the edge of the central $\beta$-sheet where the $\beta$6 strand of chain A packs anti-parallel to the $\beta$6 strand of the chain B symmetry mate. The
two sets of symmetry contacts are on nearly orthogonal axes to each other (Fig. 2-4) and may explain why despite such a high (75%) solvent content Atlpmk crystals are able to grow to nearly 1mm dimensions. Large solvent channels are evident when symmetry mates are generated, but are not large enough to fit a third monomer when the presence of disordered residues are accounted for. All disordered regions project into these solvent channels and are not near any packing interactions.

**Figure 2-4: Symmetry mate crystal packing**

Unfortunately attempts to crystallize IPMK in conjunction with any IP substrate species were unsuccessful. Kinetic tests (presented in chapter three) show the full-length construct is fully functional and the truncated region in the Δ16 Atlpmk construct is contained in the N-terminus, which is disordered in both the truncation construct and full-length seleno-methionine maps. Experimental and Δ16 2F₀Fᵋ maps agreed throughout the density, indicating that removal of the first sixteen residues had not resulted in a distortion of the structure.
Similarities and Differences to Yeast IPMK and IP3K

With a conserved ATP binding domain sequence and similar activities, \textit{AtIpmk} was expected to have a core structure highly homologous to \textit{ScIpmk} and \textit{HsIpm3k}. As expected, upon overlaying the structures of all three, the latter two solved with either ADP or AMP-PnP and IP\textsubscript{3} present, the cores of each protein were nearly identical (RMS=1.067Å and 1.118Å, Fig. 2-5).

\textbf{Figure 2-5: Structural alignment of IP kinases}

Stereo image of structural alignments of \textit{AtIpmk} (blue), \textit{ScIpmk} (green, 2IEW), and \textit{HsIpmk} (yellow, 1W2C).

Similarly the active sites of all three proteins are well conserved, with the majority of conserved residues in either the central β-sheet or secondary-structure elements that support it. The structure of the IP-binding domain is highly conserved between both IPMK species and diverges from \textit{HsIpm3k} with the additional IP-binding domain inserts (Fig. 2-6) that provide IP3K with its greater substrate selectivity (Gonzalez et al., 2004; Miller and Hurley, 2004). Comparison to the IP3K structure also confirms that like \textit{ScIpmk}, \textit{AtIpmk} has a single active site. Additional binding domains
are not observed indicating that the single active site must somehow be able to distinguish between and accommodate multiple different IP species.

**Figure 2-6: Active sites of AtIPmk and HsIP3k with IP₃**

Stereo view of the aligned active sites of AtIPmk (blue) and HsIP3k (yellow, 1W2C). I(1,4,5)P₃ is modeled in the 6-kinase orientation based on the inositol ring coordinates of the 3-kinase orientation from the substrate-bound IP3K structure. Note the more extensive packing contacts around the inositol ring due to the additional HsIPmk residues.

Uniquely Asp98, part of the PxxxDxKxG motif, was observed to directly interact with Arg152 and adopt a conformation approximately 90° offset at χ₁ from those observed in the ScIPmk and HsIP3k structures. Outside of the conserved core there are several differences from both ScIPmk and HsIP3k. As previously noted, these differences are concentrated in the external loops and species-specific inserts and in the expanded inositol binding domain of IP3K. No structural equivalent to the yeast poly-Asp region reported by Dubois et al. as required for arginine metabolism (Dubois et al., 2000) was observed. While residues 231-237 are disordered, the locations of residues 230 and 238 and the shortness of the loop indicate that AtIPmk cannot form a loop of similar size, shape, nor composition to the forty-eight residue ScIPmk insert.

Previous work in the York lab has shown that the arginine/ornithine growth defect of ipmkΔ yeast could be rescued by AtIPmk, but only if the kinase activities were intact.
(Stevenson-Paulik et al., 2002). Nevertheless the question remained if there was an equivalent region to the poly-aspartate insert of yeast IPMK also implicated in arginine metabolism. The structure of \textit{Atpmk} clearly shows that rescue is not a function of any similar structural component to the poly-aspartate insert. Not only does \textit{Atpmk} clearly lack the insert (Fig. 2-7.A), but the structure reveals that short of a complete unfolding of strands \(\beta9\) or \(\beta10\), there is no way a similar conformation could be adopted.

**Figure 2-7: Yeast poly-aspartate insert and structural comparison**

\(A.\) Structure-based sequence alignment of regions flanking yeast poly-aspartate zone. Coil (line) and helix elements are shown above and below the plant and yeast sequences. Disordered regions are shown as a dashed line. \(B.\) Disordered loops and surrounding structures. Stereo image of the plant and yeast structures around the disordered loops including the yeast poly-aspartate insert. Residues immediately before and after the disordered regions are colored in blue (plant) or green (yeast).

While residues 231-236 are disordered, any possible loop would be far too small to imitate the size of the 28-residue disordered region in yeast that includes the poly-Asp domain. Furthermore, the disordered region containing the poly-Asp domain begins pointed in the opposite direction of the disordered region in the plant structure (Fig. 2-7.B) and much further down the side of the protein; with only six residues the plant disordered region would be unable to imitate even the residues of the yeast insert that are
observable, much less the disordered region. Unfolding of the β9 and β10 strands may provide residues to form some type of loop structure, but would require disassembling part of the core of the protein and have no sequence identity to the poly-aspartate insert.

Initial Substrate Selectivity Implications

Like Holmes and Jogl with ScIpmk, I(1,4,5)P₃ from the overlaid structure of AMP-PnP and IP₃ bound HsIpk was modeled into the active site of AtIpmk. While this provides an indication of possible binding modes, by itself it does not reveal the basis of substrate selectivity in either IPMK homolog. Opposite the rat homolog (Fujii and York, 2005), plant and yeast IPMK phosphorylate IP₃ to IP₅ in a 6-/3-order (Stevenson-Paulik et al., 2002). Therefore IP₃ modeled in the 3-kinase orientation from IP3K offers little information of itself regarding substrate selectivity. What it does provide is the probable position of the target hydroxyl and ring orientation and packing within the active site, parameters that might then be applied to modeling other IP species.

**Inositol and IPMK**

Introduction

Central to investigations of IPMK and inositol signaling is the question of just how it is able to distinguish between multiple different IP species and catalyze the addition of a phosphate to the correct hydroxyl. With identical masses and electrostatic charge potentials, the only difference between different IP species with the same number of phosphate groups is the arrangement of those phosphates. Similarly unclear has been the mechanism by which IPMK identifies the correct hydroxyl to phosphorylate. The 2’-,
3’-, and 6’-hydroxyls are all unoccupied and potential phosphorylation targets. No 2-
kinase activity has been reported for any IPMK family member, not surprising as the
geometric parameters are likely quite different to add a phosphate at the axial versus
equatorial hydroxyl. An active site designed to bind inositol in the correct orientation for
one reaction is unlikely to be able to perform the other. Such dissimilarities do not exist
between the 3’- and 6’-hydroxyls. Indeed, not only are both equatorial, but IPMK will
catalyzed the phosphorylation of either position depending on the identity of the IP
species and homolog. The rat homolog adds in a 3/6-order, while the plant and yeast
forms in a 6/3, clearing indicating that some method of selection must exist. The first
step in addressing this problem is to examine the nature of inositol and its phosphorylated
species and reexamine the manner in which the question of selectivity is being asked.

The Structural Richness of Inositol

Inositol is small, simple six-carbon cyclic alcohol distinguished by the 2’-
hydroxyl existing in the axial configuration while all the remaining hydroxyls are
equatorial. However, due to the existence of one hydroxyl in the axial conformation
every inositol polyphosphate species is sterically unique. No species can be
superimposed upon the other by any combination of rotation in the x-, y-, or z-axes.
Therefore each species can potentially have unique interactions with other molecules.
This provides the basis for a large pool of signaling molecules can be easily modified by
the removal or addition of a single phosphate group. In this fashion $2^6$ different IP
species (including unphosphorylated inositol) are possible, and the number expands to $3^6$
if pyrophosphorylation is included. Without a structure of an inositol polyphosphate bound to IPMK the mechanism of selectivity has remained elusive.

Previous Selectivity Theories

Before structures were available for any of the inositol kinases it was proposed that the 1’- and 4’-phosphates bound in two conserved pockets in the IPMK binding domain (Ongusaha et al., 1998). This concept was revisited (Shears, 2004) but failed to answer several issues with the original proposal. While the 1’- and 4’-phosphates are indeed present in all reported substrate IP species except I(4,5)P₂, which is a very poor substrate (Saiardi et al., 2001), they are also present in the non-substrate IP species I(1,4)P₂ and I(1,3,4)P₃. While conservation of the 1’- and 4’-phosphates may play a role in substrate selectivity, clearly they are not sufficient in and of themselves.

More importantly, conservation of the 1’- and 4’-phosphate groups and requiring that each be bound in conserved phosphate pockets quickly results in a model that contradicts several known characteristics of IPMK. IPMK has a single active site, according to sequence analysis and the generation of kinase-dead mutants by mutating single residues (Odom et al., 2000). If the 1’- and 4’-phosphates must be bound in two specific pockets, even if the inositol is flipped such that the 1’- and 4’-phosphates switch positions, at least two different positions and orientations of the target hydroxyl must be tolerated. Using I(1,4,5)P₃ as a reference point (Fig. 2-8.A), different substrate orientations or positions of the target hydroxyl are required to fit the 1’-, 4’-phosphate pocket motif. Omitting the 5’-phosphate to reveal the hydroxyl shows a completely different position from the 3’-hydroxyl when the positions and orientations of the 1’- and
4’-phosphate groups are maintained. Flipping the entire inositol polyphosphate to reverse the positions of the 1’- and 4’-phosphates places the 5’-hydroxyl in the same position as the original 3’-hydroxyl, but now pointed in an opposite direction relative to the plane of the inositol ring (Fig. 2-8.B). Placement of the 6’-hydroxyl into the same “pocket” as the 3’-hydroxyl requires the same ring flip as when placing the 5’-hydroxyl (Fig. 2-8.C). If either the 5’- or 6’-hydroxyls are modeled with the same position and orientation of the target hydroxyl as the reference 3’-hydroxyl, it is not possible to conserve occupation of the 1’- and 4’-phosphate groups (Fig. 2-8.D). These multiple changes in substrate positioning and orientation are not consistent with the general concept that chemistry is strongly dependant on the environment and relative positions of the reacting components. While it is certainly possible to have an active site that can accommodate multiple substrate orientations, this is the exception rather than the rule. All of the phosphorylations catalyzed by IPMK are the addition of a single phosphate from ATP to a free hydroxyl on an inositol polyphosphate species performed by a single active site. It seems highly unlikely that the active site has evolved to perform the same chemistry in multiple ways. More likely is some mode of binding that ensures the participants in the chemistry are presented in a conserved manner to the catalytic site and therefore the reaction mechanism is identical for all species of IP substrates.
Figure 2-8: Comparison of equatorial hydroxyl positions in conserved 1'-4' motif

A. I(1,4,5)P₃ as reference position for 3-kinase reaction. B and C. Placement of the 6'- or 5'-hydroxyl while conserving the 1'-4'-phosphate positions requires a ring inversion. C. Ring hydroxyl positions if inositol is inverted. D. Conserving the position and orientation of the 5'-hydroxyl violates conservation of the 1', 4'-phosphate groups. The phosphate groups are numbered for the ring carbon they are attached to. The ring-position of the hydroxyl target for phosphorylation is numbered inside the inositol ring.

Chang and Majerus independently proposed a mechanism along these very lines (Chang and Majerus, 2006) for human IPMK. They proposed that the two primary substrates of human IPMK, I(1,3,4,6)P₄ and I(1,3,4,5)P₄, bind using a conserved phosphate pocket motif wherein which phosphate on the ring that occupies a position in the active site is irrelevant, only its orientation and presence. In this model the numbering of the inositol ring is of no significance, as the only factor for consideration is the presence of a hydroxyl to be phosphorylated and phosphate groups to ensure binding and recognition in the correct positions and orientations. Unfortunately the motif
proposed for HsIpmk cannot be applied to the yeast or plant homologues; it excludes a 6-
kinase activity as there is little or no such activity in the human homolog.

**An Inositol Substrate Motif**

This work was pursing the concept that substrate selection by IPMK is a structural
process indescribable by the inositol ring nomenclature at the same time that Chang and
Majerus proposed the HsIpmk substrate motif. Comparison of the larger number of IP
substrate species of yeast and plant IPMK lead to a similar but more flexible motif for
plant and yeast IPMK than that proposed by Chang and Majerus, to account for the fact
that the plant and yeast homologues are less selective and have a greater range of
activities.

To identify a possible IP-substrate recognition motif the structures of all known
substrate and non-substrate IP species were compared. One key assumption was made –
all inositol polyphosphate substrate species bind with a conserved ring and target
hydroxyl orientation. The basis of this assumption is the identical nature of all the
inositol phosphorylation reactions regardless of the identity of the target hydroxyl. Each
phosphorylation of an IP species is the addition of the ATP $\gamma$-phosphate to an unoccupied
inositol hydroxyl. With a single active site the logical expectation is that the target
inositol hydroxyl must always occupy the same position and orientation to be properly
positioned for catalysis. Similarly the positions of the inositol phosphate groups could be
expected to bind in conserved positions and orientations in a common active site.
Different binding orientations would presumably place the target hydroxyl and inositol
phosphate groups in the wrong positions to interact with the active site in a functional manner.

Two modeling parameters were used to align the various IP species following this assumption. First, the position of the target hydroxyl was conserved. Second, the ring orientation and position were also conserved. Conserving the positions and orientations of the IP species should in theory also conserve the potential interactions with IPMK. The canonical substrate of IPMK, I(1,4,5)P$_3$, was used as the initial template. Additional IP species were overlaid with the kinase-target hydroxyl modeled in the same position as the 6’-hydroxyl of the template and conserving the position and orientation of the inositol ring. IP species with multiple free hydroxyls were modeled several times, such as I(1,4,5)P$_3$ in the 3’- and 6’- orientations (Fig. 2-9.A). This process was performed for all known substrate and non-substrate IP species. Several phosphorylation patterns quickly became apparent.
Figure 2-9: Substrate and non-substrate inositol polyphosphate species

A. IP species known to be IPMK substrates. Each IP species is in the column matching the kinase target hydroxyl (circled). All species in each column are shown in the same orientation. B. IP species known to not be IPMK substrates. Non-substrate IP species are shown with various hydroxyl groups and aligned as for the substrate species. Note that the third column has IP species in three different orientations. The 1’ ring carbon and target hydroxyl are numbered and the latter circled. Phosphate groups are represented by the letter P. Not all non-substrate species orientations are shown.

As a frame of reference inositol is drawn with the kinase target (κ) hydroxyl pointed towards the bottom of the page and projecting toward the viewer. Using this
framework two characteristics are evident for all substrate IP species. The first two positions clockwise of the κ-hydroxyl are always phosphorylated. Of the first two positions counter-clockwise of the κ-hydroxyl at least one must be phosphorylated. However, no species with an axial phosphorylated hydroxyl in one of those two positions that projects down relative to the inositol ring is a substrate. The final position, opposite of the κ-hydroxyl, appears to have no role. It is both phosphorylated and not phosphorylated in substrate and non-substrate species (Fig. 2-10.A).

**Figure 2-10: Inositol polyphosphate species and proposed binding motif**

*A. Representative substrate IP species and a similar non-substrate species.* The three IP₄ substrate species. I(1,2,4,5,6)P₄, one of two non-substrate species with the same required phosphate groups. 

*B. Inositol polyphosphate binding motif.* Proposed binding motif for IP species. Required phosphate groups are in bold, partially required indicated by arrows, and exclusionary crossed out. The target hydroxyl is circled and phosphate groups are represented as P.
No non-substrate species match the first two patterns except I(1,2,4,5,6)P₅, which has an axial 2’-phosphate group in the first position counter-clockwise of the κ-hydroxyl. Often one of the two conserved phosphates is missing, such as for I(1,3,4)P₃ in the 5’- and 6’-hydroxyl orientations. The 2’-hydroxyl cannot be modeled into the same position as an equatorial hydroxyl without complete loss of coordinate conservation of the rest of the inositol ring and phosphate groups. These patterns described a binding motif (Fig. 2-10.B), presumably corresponding to the phosphate-binding pockets and surrounding residues in the active site of IPMK.

In this motif a substrate IP species must bind in such a manner that at least three pockets on IPMK are occupied by equatorial phosphate groups. The pockets are labeled α through ε moving clockwise from the κ-hydroxyl when orientated as previously described. Phosphates must occupy the α- and β-pockets and are predicted to have close interactions with the surrounding residues. At least one of the δ- and ε-pockets must also be occupied and likely have longer-range interactions with the surrounding residues. The γ-pocket is probably solvent exposed or surrounded by highly flexible residues that can accommodate a hydroxyl or phosphate group. An axial phosphate group may not be present in the α-, β-, or ε-pockets as this would likely result in either the loss of necessary interactions (α and β) or steric conflicts (ε). However an axial-phosphate would be appear to be permitted in the σ-pocket, where due to the ring geometry it would pointed in the opposite direction relative to the plane of the inositol ring than an axial group in the ε-pocket. Selection against an ε-pocket axial hydroxyl suggests that the κ-hydroxyl
points inwards towards IPMK, as an axial phosphate group doing the same would likely encounter steric clashes with the protein structure.

**Inositol Binding Modes in IPMK**

Using the overlaid coordinates of IP$_3$ from IP3K to fix the positions of the ring carbons, IP$_6$ was modeled using PyMOL (DeLano, 2002) in all six possible ring orientations (Fig. 2-11). By removing specific phosphate groups any inositol polyphosphate species could be modeled for each orientation. The interactions in each phosphate-binding pocket were identical for every IP species, as expected since the alignment was based upon conservation of the inositol ring orientation. What differs between IP species and orientation is *which* phosphate pockets are occupied.

**Figure 2-11: Inositol binding modes**
A. Substrate orientations. Inositol modeled with the 6’, 3’, or 5’-hydroxyls in the same position and orientation observed in the substrate-bound structure of HsIP3k (1W2C).  

B. Non-substrate orientations. Inositol modeled with the 1’- or 4’-hydroxyls as in A. The 2’-hydroxyl orientation was modeled conserving only the ring position and orientation as the axial hydroxyl cannot be modeled to match an equatorial hydroxyl.

The inositol binding pockets of AtIpmk

**α-pocket** – Direct interactions between AtIpmk and an IP phosphate group are observed only in the α-pocket. Hydrogen bonds are predicted between the His250 Ne2 and Tyr114 OH atoms and a phosphate oxygen.

**β-pocket** – No direct interactions were observed between in this pocket. Lys117 may adopt rotamers that directly interact, but these are of course not observed in the absence of a co-crystal structure.

**γ-pocket** – While direct interactions are not observed, Lys117 and Lys121 are in close proximity and may form direct interactions upon binding of an IP species. No other residues are in sufficient proximity to interact in any conformation. Sufficient volume exists to allow an axial phosphate group projecting toward IPMK rather than out into the solvent, as in the case of I(1,2,3,4,6)P5 in the 5-kinase orientation.

**δ-pocket** – Similar to the γ-pocket, Lys117 and Lys121 may form direct interactions in the presence of a phosphate group. If phosphates occupy both pockets Lys117 can modeled to interact only with the γ-phosphate, which blocks access to the δ-phosphate.

**ε-pocket** – Multiple possible direct contacts exist for a phosphate in the ε-pocket. Lys100, Arg133, and Arg156 as well as Lys121 may adopt interacting conformations.
Like Lys117 and the δ-pocket, a direct interaction between Lys121 and an ε-phosphate is not possible if a phosphate occupies the δ-pocket. An axial phosphate group projecting into IPMK cannot be modeled without severe steric clashes with Thr105 and Arg133. This may explain why IP species such as I(1,2,4,5,6)P₅ are not substrates, as binding would require an axial phosphate in this position.

κ (kinase-target)-pocket – Lys100 is the equivalent residue to Lys133 in the yeast homolog, identified as required in catalysis and a potential hydrogen bond to the κ-hydroxyl is observed. A simultaneous interaction with the ε-phosphate can be modeled which may play a role in catalysis through activation of lysine or stabilization of the transferring phosphate.

Figure 2-12: Residues in proximity to inositol phosphate groups

Conservation of the α- and β-phosphate groups in all known IP substrates is partially explained by the modeled interactions of these groups with IPMK. Direct interactions by less mobile residues, Tyr114 and His250, are observed with the α-
phosphate without any need to model additional conformers. Furthermore the conformation of His250 is identical to that seen for the matching histidine residue in the substrate-bound HsIP3k structure. The role of the β-phosphate is less clear, no direct interactions with the surrounding residues are observed that may explain the apparent conservation in all substrate IP species. A possible role of the β-phosphate is to sterically restrict the ability of the α-phosphate rotate about the oxygen-phosphate dihedral bond such that a conformation interacting with Tyr114 and His250 is preferred. In the absence of a β-phosphate the α-phosphate may adopt a conformation that prevents binding or is unable to properly coordinate elements of the active-site.

The requirement for either the δ- or ε-pockets to be occupied by at least one phosphate group may be explained by the numerous potential interactions with several long-chain residues. Several arginine and lysine residues appear capable of direct interactions with a phosphate group occupying either of these pockets. These interactions could provide additional binding affinity and allow greater substrate flexibility by accommodating an IP species as long as it has a phosphate in one of the two pockets. A lack of selectivity at the γ-position is likely due to the few potential interactions and primarily solvent-exposed nature of phosphate at this position. With few interactions possible there is likely to be little if any binding contribution by a phosphate group and therefore no requirement for one. Similarly no exclusionary affects are predicted even if an axial phosphate group, which would project toward the IP binding domain, is present. There is sufficient volume open around the γ-phosphate that no steric conflicts are modeled and therefore the presence of a phosphate does not play an exclusionary role.
Conversely the inability to accommodate an axial phosphate projecting toward the core of IPMK in the ε-pocket explains why certain species that fulfill the equatorial phosphate requirements are not substrates. While the coordinating phosphate groups are present, the axial phosphate would prevent binding by steric blockage.

A 1’-, 4’-phosphate binding pocket is not supported

Ongusaha proposed and Shears reviewed the possibility that at least part of IPMK selectivity was based on a conserved binding motif where the 1’- and 4’-phosphates always occupied the same pockets. The multiple orientations required to maintain the target hydroxyl in the same approximate position or multiple positions if the inositol ring coordinates were conserved argue against this hypothesis (Fig 2-8). Examination of the AtIpmk structure provides further evidence that selectivity is through a different mechanism. As previously described, the coordinates for IP3 from the overlaid structure of HsIp3k, IP3 was modeled into active site of AtIpmk, placing IP3 in the 3-kinase orientation. If the 1’- and 4’-phosphate pockets are defined by the positions of these two groups for the 3-kinase orientation they occupy the α- and δ-pockets. If inositol is modeled in the 6-kinase orientation the β- and ε-pockets are occupied. Conserving either set of pockets does not result in a binding model consistent with the structure of AtIpmk.

In the 3-kinase orientation phosphate groups are conserved at the α-site, where the most interactions with AtIpmk are predicted, and the δ-site, where interactions are predicted to be with the highly mobile residues Lys121 and Arg156. If these pockets are supposed to be highly conserved, it is unlikely that one of them would be made of residues with multiple possible non-binding conformations. More significantly, inversion
of the inositol ring to bring the 6’- or 5’-hydroxyls into the \( \kappa \)-pocket while maintaining occupation of the \( \alpha \)- and \( \delta \)-pockets places both phosphate groups in different positions. While the difference may not be significant in the \( \delta \)-pocket, where Lys121 and Arg156 might be able to adopt binding conformations, this is not the case in the \( \alpha \)-pocket. If the geometry of the phosphate group is flipped, the oxygen atoms of the phosphate group are no longer close enough to interact with Tyr114 or His250. Moving the phosphate group into the same position as seen for the 3-kinase orientation shifts the rest of the inositol polyphosphate to such a degree that the target hydroxyl is now out of position and steric conflicts occur for the \( \delta \)-phosphate. Additionally, \( I(1,4)P_2 \) is not an IPMK substrate – additional phosphate groups would be present and when modeled, such as for \( I(1,4,5)P_3 \), these groups have either disrupted interactions with \( A/Ipmk \) or come into steric conflict once the ring is inverted.

Conservation of the 1’-, 4’-phosphates is also not supported in the 6-kinase orientation. The \( \beta \)- and \( \varepsilon \)-pockets are occupied in this model and encounter similar issues to the 3-kinase orientation. No direct interactions are observed for the \( \beta \)-phosphate for any modeled IP, directly contradictory to the concept that a phosphate must be present for successful binding and catalysis. The motif proposed in this work suggests that the \( \beta \)-phosphate is essential not for direct interactions with the protein, but to help maintain the correct position of the \( \alpha \)-phosphate. A similar argument could be made for the \( \beta \)-phosphate in the 1’-, 4’-pocket motif, but then a third phosphate group would be conserved, which is not observed in the substrate species if the positions of the 1’- and 4’-phosphates must be conserved. Occupation of the \( \varepsilon \)-pocket has a similar concern to
the δ-pocket in the 3-kinase orientation; potential contacts are with highly mobile residues rather than a tightly conserved structure. Flipping or rotating various IP species to place the 3’-hydroxyl in the active site again encounters issues with position of the target hydroxyl and phosphate group interactions. If flipped to maintain phosphate groups in the β- and ε-pockets, the target hydroxyl position and orientation are shifted. Moving the inositol ring to bring the target hydroxyl into the same position occupied by the 6’-hydroxyl either brings the phosphate groups into steric conflict with the binding site or requires such a rotation of the molecule that no interactions exist between the phosphate groups and the binding site. The 5-kinase orientation can be modeled to occupy the same pockets with the same binding geometry, resulting in the same bound models as seen for the motif proposed in this work.

Avoiding the problem of ring inversion is possible by simply maintaining the ring coordinates. However, the structure of Atlpmk also argues against this model. If the α- and δ- or β- and ε-pockets are conserved and the 1’- and 4’-phosphates must always occupy these pockets in the same ring orientation, multiple different target hydroxyl positions are required (Fig. 2-8). Examining the structure of Atlpmk indicates no tolerance for such different hydroxyl targets. There is clearly no binding site for ATP on the opposite side of the inositol ring to phosphorylate the 5’-hydroxyl (Fig. 2-8.A, if the 5’-phosphate were missing), nor for the 6’-hydroxyl completely opposite of the 3’-hydroxyl. If modeled as in Fig. 2-8.C with the conserved phosphate groups in the β- and ε-pockets, the 3’-hydroxyl is now pointing toward part of the IP binding domain, where there is clearly no room for an ATP nor any catalytic residues. The structure of Atlpmk
confirms that it does not have a second binding or active site and therefore does not support an IP binding model that requires multiple target hydroxyl positions, especially ones on opposite sides of the inositol ring.

Clearly the inositol polyphosphate selectivity of \( A\)Ipmk cannot be explained by conservation of the 1’- and 4’-phosphates in specific pockets. Indeed, any motif that describes the conserved phosphate groups by their numerical identification on the inositol ring encounters the same difficulties. While the 1’- and 4’-phosphates are present in all substrates, they clearly cannot occupy the same positions in the binding site in all substrates. Rather they must occupy two of several different phosphate binding pockets as proposed in this work.

Why \( A\)Ipmk is not a 1/4-kinase and specificity from steric exclusion

The lack of a 1/4-kinase activity may be explained by the models of various IP species in the IPMK active site. All known substrate species when modeled into the active site have equatorial phosphates in both pockets. The \( \alpha\)-phosphate forms hydrogen bonds with Tyr114 and His250, as previously discussed and shown in Figure 2-13.A. No IP species with the 1’- or 4’-hydroxyl in the \( \kappa\)-pocket has equatorial phosphates occupying both the \( \alpha\)- and \( \beta\)-pockets. In the 1’-orientation the \( \alpha\)-pocket would instead be occupied by the 2’-hydroxyl or an attached phosphate group. With no phosphate group the IP molecule likely lacks interactions necessary for binding affinity and coordination of the active site. A 2’-phosphate would not be accommodated within the pocket, instead coming into steric conflict with Tyr114 and Ser219 (Fig. 2-13.B). Accommodation of a 2’-phosphate in the \( \alpha\)-pocket would require shifting the entire IP molecule up and away
from the binding site, likely breaking multiple interactions as well as removing the target hydroxyl from proximity to the active site residues and ATP. An IP species such as I(1,3,5,6)P$_4$ might be able to bind in the 4’-kinase orientation, but the β-pocket would not be occupied. In the 4-kinase orientation the β-position is occupied by the 2’-axial hydroxyl group which cannot form any direct interactions with the surrounding residues, even if a phosphate group is present. Due to axial nature of the 2’-hydroxyl an attached phosphate group will project out into the solvent region and away from any possible interactions with IPMK. Not only are no 2’-phosphate interactions with AtIpmk possible, but it will also be unable to stabilize interaction of the 3’-phosphate in the α-pocket (Fig. 2-13.C).

Not all IP species with α- and β-phosphates are substrates, as seen in the case of I(1,2,4,5,6)P$_5$. Any ε-phosphate in a substrate IP species is equatorial in all models (Fig. 2-14.A). However, when I(1,2,4,5,6)P$_5$ is modeled in the 3-kinase orientation the 2’-phosphate projects down and into the core of the surrounding protein (Fig. 2-14.B). Here it comes into steric conflict with the $C\gamma_2$ of Thr105 and the $C\beta$ and $C\gamma$ atoms of Lys121. Neither residue can be modeled in conflict-relieving conformations. Presumably this steric conflict prevents any IP species with a 2’-phosphate group from binding in the 3-kinase orientation. The presence of a 2’-phosphate does not, however, prevent an IP species from being a substrate for IPMK. I(1,2,3,4,6)P$_5$ is phosphorylated at the 5’-hydroxyl to form IP$_6$, clearly indicating that IP binding must occur in a mode that does not bring the 2’-phosphate into steric conflict with the surrounding residues. When I(1,2,3,4,6)P$_5$ is modeled in the 5-kinase orientation the 2’-phosphate is in the γ-pocket,
projecting toward the protein core. However, the 2’-phosphate does not play an exclusionary role. Unlike when the 2’-phosphate is in the ε-pocket, there is sufficient volume below the γ-pocket to accommodate a 2’-phosphate group. It is in fact more accurate to state that the α-ε pockets, being defined as the volumes of space occupied by the equatorial phosphate groups, can never be occupied by an axial phosphate group. Binding of IP species with 2’-phosphate groups then requires examination of the different volumes that would be filled and any potential interactions. This raises a new question – if an IP species is predicted to occupy the α- and β-pockets, but has only an axial phosphate in the δ-pocket and no equatorial phosphate in the ε-pocket, will it be a substrate? If the ε-phosphate is now present in a different IP species, does that IP become a substrate? IP species that might fit these parameters, I(2,4,5)P₃ and I(1,2,4,5)P₄ have not been reported to have been tested for activity, and indeed the latter is not even commercially available.
Figure 2-13: Equatorial phosphate requirement in α- and β-pockets

A. Inositol in the 3-kinase orientation with equatorial phosphates in the α- and β-pockets. Interactions of the surrounding residues with inositol and the phosphate groups are shown as contact dots. Residues Y114 and H250 form hydrogen bonds to the 4-phosphate in the α-pocket while there are no 5’-phosphate interactions in the β-pocket. Rotational freedom of the α-phosphate about the C-O bond is restricted by the presence of the β-phosphate, which may stabilize the α-phosphate interactions.

B. 1-kinase activity is prevented by loss of interactions or steric conflict. Inositol is modeled in the 1-kinase orientation with a 2’-phosphate. In the absence of a phosphate group all interactions with Y114 and H250 are lost. When a 2’-phosphate is present an extreme degree of steric conflict occurs with Y114 and H250, presumably excluding any such IP species from the binding site and therefore preventing a 1-kinase activity.

C. 4-kinase activity prevented by loss of packing. When inositol is modeled in the 4-kinase orientation the packing interactions between the α- and β-phosphates is lost due to the 2’-axial hydroxyl. Even with a phosphate present there are no packing interactions between the 2’- and 3’-phosphates that would stabilize the interaction. While hydrogen-bonding between the 3’-phosphate to Y114 and H250 is possible, there also exists significant rotational freedom about the C3-O3 bond. See the appendix for a description of contact coloring and generation method.
Figure 2-14: 2'-phosphate exclusion from the ε-pocket

A. An equatorial phosphate in the ε-pocket has multiple binding partners. I(1,3,4,6)P$_4$ is shown in the 5-kinase orientation with the 4'-phosphate occupying the ε-pocket. Hydrogen bonds are shown to K100, R133, and R156 and there are no steric conflicts. 

B. A 2'-phosphate is excluded from the ε-pocket. I(1,2,4,5,6)P$_5$ is shown in the 3-kinase orientation with the 2'-phosphate in the ε-position. The phosphate group is actually below the ε-pocket and protrudes towards the core of Atpmk where it comes into steric conflict with K100 and K121. For clarity only steric conflict interactions are shown. The only difference between the non-substrate I(1,2,4,5,6)P$_5$ and substrate I(1,4,5,6)P$_4$ species is the presence of the 2'-phosphate.

All known substrate and non-substrate IP species match the equatorial and axial phosphate criteria. From these models the mechanism of IP recognition by IPMK may best be described as using a topological surface to screen for appropriate substrates.

Rather than IPMK recognizing a specific set of IP species, similar to how IP3K will bind only I(1,4,5)P$_3$ and in only the 3-kinase orientation, it instead phosphorylates any IP species that can fit into the binding domain. The lack of 1/4-kinase activity is not due to some specific recognition that such IP species are not substrates, but is instead likely a failure of these species to bind or bind in a manner that allows phosphorylation to occur.
Inositol polyphosphate substrate recognition by *At*Ipmk

*At*Ipmk is proposed to recognize substrate inositol polyphosphate species through a recognition face that requires the presence of phosphate groups in specific pockets while excluding those with certain axial phosphates. The α- and β-pockets must be occupied by equatorial phosphates and at least of the δ- or ε-pockets must also be occupied. An axial phosphate at the 2’-position can be tolerated so long as it would not bind in the α- or ε-pockets where it would come into steric conflict with multiple side-chains. In all IP species the orientation and position of the inositol ring must be conserved, which in turn conserves the positions and interactions of the attached phosphate groups regardless of the identity of the target hydroxyl for phosphorylation.

Interactions of *At*Ipmk with the various inositol phosphate groups modeled following the proposed binding motif support this hypothesis. The most conserved phosphate groups interact with the most sterically restricted residues, phosphates that are less strictly required have multiple possible interactions with highly flexible residues, and IP species reported to have no activity yet fulfilling the equatorial phosphate requirements also have axial phosphates predicted to prevent binding. All known substrate and non-substrate IP species fit this model.

**Summary**

As predicted the structures, *At*Ipmk, *Sc*Ipmk, and *Hs*Ip3k, have similar cores while external regions differ. Elements involved in the common kinase properties of each enzyme would logically be expected to have similar structures – performing the same reactions upon the same substrates, a similar structure would be expected and is
indeed observed. As was noted by Holmes and Jogl for the yeast homolog, the inositol polyphosphate binding domain of plant IPMK is much smaller than is seen for the more specific IP3K. The reduced number of potential contacts with the various inositol phosphate groups is responsible for the greater substrate variability of the IPMK species. When other IP species are modeled into the active site of IP3K or IP3 is modeled with different hydroxyls in the position of the target 3’-hydroxyl, multiple steric clashes occur between the IP phosphate groups and the residues unique to IP3K. Lacking these extra residues the IPMK binding site can accommodate more IP species and orientations.

The structure of AtIpmk also clearly shows that it does not possess a domain similar in structure to the poly-aspartate region of ScIpmk. This further supports the hypothesis that IP production alone is sufficient to rescue arginine/ornithine metabolism in ipmkΔ yeast. Comparison with ScIpmk and HsIpmk has also identified a possible target for additional investigation of the signaling roles of ScIpmk based on protein-protein complex formation, such as the ArgR-Mcm1 complex. The yeast residues Lys174-Ser193 form an additional small helix and strand at the edge of the central β-sheet that is unique to the yeast homolog. No reports exist of any examination of this region, but further investigation may be warranted as it appears to be a unique feature of the yeast homolog.

While the structure of AtIpmk provides some interesting insights, the lack of a co-crystal structure with an IP species means it was not possible to directly observe the substrate selectivity mechanism. Efforts turned to an examination of the inositol polyphosphate species identified as substrate or non-substrate to look for a
phosphorylation motif. It quickly became apparent that trying to identify a motif based on the numbering of the inositol ring would not lead to an answer, as there are no phosphates unique to either substrate or non-substrate species. Instead the structural similarities and differences of each species were compared using a common frame of reference. By aligning every species to conserve the ring orientation and position of the target hydroxyl for phosphorylation a phosphorylation motif quickly became apparent. This motif is independent of the inositol ring numbering, which itself is simply a descriptive tool.

Modeling IP species into the structure of AtIpmk reveals that the substrate binding motif matches extremely well to the surrounding residues. Phosphate groups predicted to be absolutely required had the closest and most potential interactions with IPMK (the α-phosphate) or reinforced those phosphates (the β-phosphate). Those less required could interact with the same residues and therefore fulfill a binding requirement as long as at least one was present (the δ- and ε-pockets), and the apparently non-selective γ-pocket was almost entirely solvent exposed and could accommodate any phosphorylation state. Similarly the inability of IPMK to phosphorylate several IP species is explained by either a lack of the required phosphate groups, or an inability to bind in the first place as the axial phosphate on several species was modeled to cause severe steric conflicts with the core of the protein.

From the structures of the various inositol polyphosphate species is proposed a recognition motif that addresses all currently known substrate and non-substrate species. Modeling IP species into the structure of AtIpmk following this motif reveals no
contradicting information and provides further support through the steric clashes of disallowed species with the protein backbone.

**Experimental Methods**

*Modeling of IP substrates into the active site of AtIpmk* – The structures of AtIpmk, ADP-bound ScIpmk (2IF8), and substrate and product bound HsIp3k (1W2C and 1W2D) were overlaid with LSQMAN (Kleywegt, 1996). The coordinates of I(1,4,5)P₃ and AMP-PnP from HsIp3k were overlaid onto the AtIpmk structure. Different IP binding modes were modeled using PyMOL (DeLano, 2002) to overlay IP₆ from the structure of ADAR2 (adenosine deaminase acting on RNA) (1ZY7) upon the inositol ring orientation of I(1,4,5)P₃. IP₆ was rotated as required to place each hydroxyl in the same position and orientation as the 3’-hydroxyl and conserve the ring-carbon coordinates. The 2’-phosphate could not be modeled in an equivalent position due to its 2’-axial configuration and was modeled by preserving the inositol ring coordinates. A library of binding modes was generated with IP₆ modeled in six different orientations, one for each ring hydroxyl matching the IP₃ 3’-hydroxyl coordinates. Phosphate groups were deleted from IP₆ to model specific IP species in each orientation.
Chapter 3. Redesigning IPMK

Introduction

While computational modeling strongly supports the proposed inositol polyphosphate selectivity hypothesis, it does not provide any form of physical support. The clear manner to test this hypothesis is to use it as a guide to design mutants that would in theory have altered substrate specificities and measure changes in inositol polyphosphate production and kinetic activity. In conjunction with the objective of creating tools able to probe inositol signaling in vivo the most efficient approach would be to design mutants with reduced or eliminated 3-kinase activity upon I(1,4,5,6)P$_4$. Recall that in yeast, plant, and flies, IPMK catalyzes the phosphorylation of I(1,4,5)P$_3$ to I(1,4,5,6)P$_4$ to I(1,3,4,5,6)P$_5$ through a 6-/3-kinase ordered reaction (Fujii and York, 2005; Seeds et al., 2004; Stevenson-Paulik et al., 2002). To expand understanding of the roles played by I(1,4,5,6)P$_4$ versus I(1,3,4,5,6)P$_5$ a construct would be needed that retained the first step but could no longer perform the second. Altering the 5-kinase reaction for I(1,3,4,6)P$_4$ would also be a potential target, to investigate the role of the human homolog which primarily performs this reaction and the 3’-phosphorylation of I(1,4,5,6)P$_4$, but is reported to have poor 6-kinase activity upon both I(1,4,5)P$_3$ and I(1,3,4,5)P$_4$ (Nalaskowski et al., 2002). Hence a redesigned IPMK would serve two purposes; first it would either support or disprove the selectivity hypothesis and second, if successful, would provide new, unique tools for investigating inositol signaling in vivo.
If an I(1,4,5)P$_3$ 6-kinase could be designed with no activity for any other IP species at any other hydroxyl, the result would be an entirely unique enzyme, the first known example of an IP$_3$-6K.

**Rational Design and the Inositol Binding Site**

Possessing an explicit model of inositol binding and structure of IPMK, a rational design process was used to propose potentially selective mutants. Rational design techniques use pre-existing knowledge to propose specific alterations for specific effects, in contrast to directed evolution where a library of randomly or semi-randomly generated mutants are screened for a desired phenotype. In the situation of designing IPMK constructs with preferences for specific inositol polyphosphate species and specific phosphate additions, rational design was considered the more efficient approach. A significant limitation to a directed evolution approach in the case of IPMK is the manner of screening for specific phenotypes that would indicate an altered kinase specificity. The primary selective conditions known for IPMK compromised cells are the failures to grow on arginine/ornithine media as the sole source of nitrogen and temperature sensitivity. While mutants could certainly be screened for survival, the ability to grow under these conditions would not indicate if there had been any change in inositol production, only that it had not been lost entirely. More importantly, identifying if there had been a change in the inositol profiles as an indicator that IP substrate specificity had been altered would require large-scale radioactive labeling assays. Each mutant would have to be grown in the presence of $[^3]$H-D-	extit{myo}-inositol, the culture processed, and separated by HPLC to measure the soluble IP populations and search for any changes.
from wild-type. The logistical and financial costs of such a screening process were considered inefficient.

Similarly it was not felt that an in vitro directed evolution technique would be the most efficient approach to redesign the IP binding site of IPMK. Since IPMK would have to be generated to test the kinase functions, an in vitro approach would require some type of cell-free system. Such a system would have to be designed and optimized for IPMK and still fail to address the fundamental question of the cost of screening a large library of constructs with a radioactive assay. It would be possible to use cold IP substrates and $[^{32}P]$-labeled ATP, reducing the cost, but still require a larger amount of resources than necessary.

In contrast a rational design methodology that used in silico methods to identify potential candidates offered several advantages. With the structure of both IP species of interest and the kinase in question known, much of the ambiguity that might exist during an attempt to redesign a substrate binding site is not present. Furthermore the design goal is to remove preexisting activity, rather than generate a novel activity, such as the ability to catalyze the phosphorylation of a previously non-substrate IP species. Most significantly the potential pool of candidates is much smaller. Using the structure and proposed binding orientations of various IP species, clearly only a few residues are likely candidates for mutation, and of those, only a few mutations are likely to have the desired effect. If the goal is to prevent an IP species from binding by blocking one of the pockets a phosphate group would occupy, clearly mutating a residue such as lysine to an alanine or glycine would not have the desired effect. A great advantage of the directed evolution
technique, the ability to screen an extremely large library of constructs, is simply not necessary to determine if the inositol binding motif hypothesis is correct.

The goal of the design process was to identify mutants with the potential to block the binding of specific inositol polyphosphate species in specific orientations to reduce the IP substrate profile. The only difference between the various IP species would be the presence or absence of a phosphate group occupying a specific position in the binding site. Therefore the design target was to identify mutations at specific residues that would occupy those volumes and block substrate binding by steric conflicts. Electrostatic repulsion was also considered, but as will be addressed was not found to offer any likely candidates.

When \(\text{I}(1,3,4,5)\text{P}_4\) and \(\text{I}(1,4,5,6)\text{P}_4\) are modeled into the active site of \(A\text{tIpmk}\) in the proposed 6- and 3-kinase orientations (Fig. 3-1.A and B) a difference in the volumes occupied by the phosphate groups is observed. In addition to occupying the required \(\alpha-\) and \(\beta-\)pockets, \(\text{I}(1,3,4,5)\text{P}_4\) also has phosphates in the \(\gamma-\) and \(\varepsilon-\)pockets. Conversely \(\text{I}(1,4,5,6)\text{P}_4\) occupies the \(\gamma-\) and \(\delta-\)pockets.
This immediately suggests the potential for two designs. One, wherein the γ-pocket were to be blocked, would be expected to inhibit the binding of either IP$_4$ species. Modeling I(1,4,5)P$_3$ in the 6-kinase orientation reveals that only the ε-pocket is occupied (in addition to the required α- and β-pockets) and therefore IP$_3$ would remain a substrate. Secondly, the two IP$_4$ species appear to be recognized in different fashions and this difference might be exploited to design a construct that favors one species over the other. A mutation that occupies the δ-pocket would be predicted to prevent binding of I(1,4,5,6)P$_4$ in the 3-kinase orientation, but allow I(1,3,4,5)P$_4$ to bind with little or no disruption. IP$_3$ would be expected to be unaffected as well.

Not surprisingly there are few residues in proximity to the proposed γ-pocket. The IP substrate binding hypothesis suggests that in wild-type enzyme the presence of a phosphate in the γ-pocket plays no role in substrate selectivity. Species with an equatorial or axial phosphate in the γ-pocket had no steric conflicts with any part of the surrounding model, further support that normally interactions in this region have no
exclusionary role. The only residue in any proximity is Lys117, in the middle of helix α3. Lys121 was identified as the residue in closest proximity and only one that might be mutated to block the volume occupied by the phosphate filling the δ-pocket. Flanking residues for Lys117 and Lys121 were too far away from the γ- or δ-pockets and as part of helix α3 were orientated such that any mutants would not be able to adopt a conformation that blocked either pocket. Using the inositol binding hypothesis in conjunction with the structure of AtIPmk only these two residues were identified as candidates for mutagenesis to generate selective constructs.

**Redesigning the IP Binding Site of IPMK with KiNG**

Entropic and thermodynamic principles state that a protein will adopt the lowest possible energy state for a given environment. In the case of protein side-chains, these states are seen as the most highly populated conformations, i.e.; rotamers. Constructs designed without taking into account the likelihood that a side-chain would adopt the necessary conformation(s) would be unlikely to exhibit the desired changes in selectivity. Therefore an *in silico* screening process using a rotamer library provided a method to test if a mutant could be reasonably expected to adopt a conformation that would increase substrate selectivity.

*In silico* screening was performed by mutating the residues Lys117 and Lys121 to residues of sufficient size to have to potential to block the γ- and δ-phosphate binding pockets. Models of AtIPmk with IP₃, I(1,4,5,6)P₄, or I(1,3,4,5)P₄ were converted to kinemage files using the MolProbity server (Lovell et al., 2003). Lys117 or Lys121 were mutated to asparagine, aspartate, arginine, glutamate, glutamine, histidine, phenylalanine,
tryptophan, and tyrosine using the KiNG program (Chen et al., 2009). Each rotamer (Davis et al., 2004) for every mutant was then examined for steric conflicts (clashes) with the surrounding residues and inositol polyphosphates. These rotamers were then examined for conformations that would be predicted to block a specific IP species from binding. However, simply containing a rotamer predicted to block an IP species (visualized by steric clashes against the modeled phosphate group) could not be expected to indicate selectivity. If a conformation existed that had no steric conflicts with the IP or backbone of IPMK it would not be expected to be selective, as in the presence of the IP species it could simply adopt the non-clashing conformation and allow IP binding. Therefore a more stringent rule was applied to screen each potential mutant.

To be a candidate for in vitro testing the model for each mutant could contain only rotamers where those without steric conflicts with the main-chain of IPMK occupied the desired phosphate pocket. Rotamers outside of the target pocket and without steric conflicts would allow IP species to bind to the mutant. Side-chain to side-chain conflicts were initially treated as no-conflict cases, to minimize the modeling complexity. This would eliminate more potential candidates and raise the possibility of selective mutants being missed, but prevent the necessity of complex modeling procedures to alter the coordinates of multiple side-chains simultaneously. Each rotamer was also rotated about each component \( \chi \) angle within the “excellent” range (Lovell et al., 2003) to ensure that the conformational space occupied by each rotamer had been examined. While a rotamer library contains specific conformations for each residue, these are only the most commonly observed within a range. A residue with a conformation within that observed
range may still be expected to exist and must also be examined for steric conflicts or a lack thereof that would allow substrate binding to proceed. Conversely the main-chain atom coordinates were not altered. Rotamers that would require altering the conformation of the entire inositol binding domain to eliminate steric conflicts were not addressed. Any rotamer requiring such significant alterations to the binding site would likely eliminate or greatly reduce affinity for any IP species by disrupting the entire binding domain. Each rotamer model was examined in the KiNG (Chen et al., 2009) program to identify the nature of any steric conflicts; inositol ring, phosphate, main-chain, or side-chain. Side-chain conflicts were examined in more detail to determine if such rotamers were likely, based on the required displacement of the conflicting side-chain(s). These parameters were selected to make the first selectivity search as rapid and restrictive as possible. Should no candidates be identified steps could be taken to reduce restrictions, but at the cost of additional computational time and more complex modeling techniques.
Figure 3-2: Design of the K121W Atpmk selective mutant

A. \textit{I}(1,4,5)P_3 \textit{in the 6-kinase orientation with K121W}. IP_3 modeled in the 6-kinase orientation and surrounding residues. Note the lack of steric conflicts with K121W. K121W is modeled in one of the two conformations that do not have steric clashes with the main-chain. 

B. \textit{Prohibitive side-chain conflicts between K121W and R133, R156, and IP_3}. In the second conformation without main-chain conflicts K121W has severe steric conflicts with both R133 and R156 as well as the 1’-phosphate (occupying the ε-pocket). This conformation is not predicted to occur or diminish selectivity as it would require displacement of two side-chains, still block the δ-pocket, and additionally block the ε-pocket. 

C. \textit{K121W is predicted to bind I}(1,3,4,5)P_4. No steric conflicts between K121W and I(1,3,4,5)P_4 in the 6-kinase orientation are apparent. 

D. \textit{K121W blocks I}(1,4,5,6)P_4. In contrast I(1,4,5,6)P_4 in the 3-kinase orientation is expected to be excluded by steric conflict in the δ-pocket between the 1’-phosphate group and K121W. No conflict-relieving conformations exist within rotameric space.
Two selective candidates were identified. K117W and K121W were predicted to block the \(\gamma\)- and \(\delta\)-pockets, respectively. No conformations of aspartate, asparagine, glutamate, or glutamine were found to even approach the phosphate pocket. These residues were simply too short to approach close enough for either steric or electrostatic repulsion. Additionally conformations existed for each that had no steric conflicts with any modeled IP species nor any part of IPMK. Therefore no selectivity would be predicted as not only were there no blocking conformations, but there were also conformations that would have allowed substrate binding regardless. Neither histidine, phenylalanine, nor tyrosine could be modeled in a conformation blocking the phosphate pockets. Not surprisingly arginine could be modeled in several conformations that would block binding of a phosphate group. Also not surprisingly multiple conformations existed that would allow binding. With four \(\chi\)-angles arginine has a high degree of conformational flexibility that allows it to adopt a large number of conformations.

Two K117W rotamers have steric conflicts with only IP\(_4\). Rotamer t90 projects into and beyond the \(\gamma\)-pocket to such an extent that it has conflicts with both the phosphate group and the inositol ring itself. Clearly it would not be predicted to provide specificity against any IP species with a \(\gamma\)-phosphate (phosphates occupying a pocket are named for the pocket as the ring position with different species) as any IP species would likely be excluded. In contrast rotamer t-105 clashes primarily with the \(\gamma\)-phosphate. Some clashes do exist between the H\(\zeta\)2 proton and C3 and O13 atoms of inositol which can be relieved by altering the \(\chi_1\) angle by approximately 20°, which still falls within the “excellent” rotamer range. Of the remaining rotamers only m0 does not have extensive
main-chain clashes that cannot be relieved by \( \chi \)-angle alterations. While there is a steric conflict between the C\( \delta 1 \) of K117W and carboxyl oxygen of Glu113, it can be relieved with a 10° shift in \( \chi_1 \). K117W would therefore be expected to have a reduced affinity for all IP species and a greater reduction for any with a γ-phosphate such as I(1,3,4,5)P\(_4\) or I(1,4,5,6)P\(_4\), but not an elimination of IP\(_4\) phosphorylation.

Figure 3-3: Predicted K117W interactions with I(1,3,4,5)P\(_4\) and I(1,4,5,6)P\(_4\)

A. I(1,3,4,5)P\(_4\) excluded by steric conflicts with K117W. The 4'-phosphate is blocked by steric conflict with K117W in the γ-pocket. Steric conflict is more severe than displayed, as the atoms are so close contact analysis fails to properly calculate all of the conflicts. B. I(1,4,5,6)P\(_4\) is also excluded by K117W. Exclusion is similar to I(1,3,4,5)P\(_4\), with conflict again against K117W in the γ-pocket. Here conflict is between K117W and the 6'-phosphate.

Analysis of K121W contacts with I(1,3,4,5)P\(_4\) and I(1,4,5,6)P\(_4\) indicated the potential for a mutant with a differential selectivity between IP species. The m95 rotamer with 15° and 10° changes to the \( \chi_1 \) and \( \chi_2 \) angles (still within the “excellent” range) had no steric conflicts with IPMK and a slight clash with the γ-phosphate of I(1,3,4,5)P\(_4\). When modeled with I(1,4,5,6)P\(_4\) extensive clashes were observed with the δ-phosphate,
which is not present for I(1,3,4,5)P₄ or I(1,4,5)P₃. Of the remaining rotamers p-90, p90 (similarly named but different), and m-90 had extensive, non-relievable main-chain clashes. Clashes with both the γ-phosphate and inositol ring were observed for the m0 rotamer. However, rotamers t-105 and t90 had clashes only with the side-chains of residues Arg133 and Arg156 or Lys100, Asp122, Arg133, and Arg156. Normally clashes of a mutant rotamer with only side-chain residues would be considered a non-clashing event and the mutant discarded from the list of candidates. However, closer examination suggested that the t-105 and t90 rotamers would not provide a functional binding mode for I(1,3,4,5)P₄, or any IP species.

Both Arg133 and Arg156 appear to be involved in interactions with the ε-phosphate, for I(1,3,4,5)P₄ the 1’-phosphate group. During analysis of the various substrate and non-substrate IP species it was observed that at least one phosphate must occupy the δ- or ε-pockets. For both IP₄ species the only residues close enough to interact with a phosphate in either pocket are Arg133 and Arg156. If K121W were to adopt either rotamer then neither arginine residue would be able to adopt a conformation that interacts with the IP phosphates, presumably the same functional effect as if no phosphate were present at all in either pocket. The tryptophan mutant itself would not be expected to form binding interactions due to its distance from the phosphate groups and lack of atoms that could form hydrogen bonds. Arguing against the possibility that K121W could even adopt the t-105 and t90 rotamers is the highly unlikely presumption that it would be more energetically favorable for one residue to directly disrupt the conformations of two to four other residues than adopt a non-conflicting conformation.
In addition secondary disruptions would be required to accommodate the new positions of the displaced residues, another penalty against adoption of the t-105 and t90 rotamers.

From these examinations K117W was predicted to be partially selective against I(1,4,5,6)P₄ and I(1,3,4,5)P₄ due to blockage of the γ-pocket. I(1,3,4,5)P₄ was also predicted to be slightly less affected than I(1,4,5,6)P₄ as the lack of δ-phosphate might allow the γ-phosphate more room to sample clash-relieving conformations. The K121W mutant was predicted to be strongly selective against I(1,4,5,6)P₄ but only slightly against I(1,3,4,5)P₄ due to strong clashes with the δ-phosphate and slight ones against the γ-phosphate. A minor loss of IP₃ activity was predicted, as general loss of affinity for all IP species due to the existence of rotamers for both mutants that would clash with the inositol ring.
Figure 3-4: Schematic of IP₄ species and Selective Mutant Interactions

Testing the Motif Hypothesis

Mutant production of IP₄ and identification of the K121W Atlpmk product

PCR mutagenesis was used to generate K117W and K121W constructs from the wild-type Atlpmk pET-15b construct and protein expressed and purified by affinity chromatography. The column-digest double-column technique was not used in this case.

Figure 3-5: K117W and K121W Atlpmk purification

12% SDS-PAGE gels of mutant production, see appendix for description.

There was a loss of solubility in both constructs, with protein precipitating out of solution when concentrated above approximately 10mg/ml. This was not unexpected as both lysine residues are solvent exposed and therefore replacement with tryptophan replaces a charged, highly mobile side-chain with a much more restricted and highly hydrophobic group that is likely to decrease solubility. The selectivity of the K117W and K121W constructs was screened by treating I(1,4,5)P₃ with [³²γ]-ATP and separating the IP products by thin-layer chromatography. Reactions were run 70 minutes and in parallel with wild-type Atlpmk, with samples taken at five minutes and every ten minutes after initiation. An IP₄/IP₅ standard and a no-enzyme control were included. Wild-type
Atlpmk generated IP₄ and IP₅, K117W generated IP₄ and significantly reduced levels of IP₅, and K121W generated IP₄ and barely detectable levels of IP₅ (Fig. 3-4). These production patterns match the predictions based on the modeled IP species and mutants.

![Image](image.png)

**Figure 3-6: Screen of IP₄ and IP₅ Production by Atlpmk and constructs**

TLC separation of IP₄ and IP₅ products of IP₃ phosphorylation with [³²γ]-ATP. Time-points were removed at 5, 10, 20, and 30 minutes.

The preference of K121W Atlpmk for IP₃ over I(1,4,5,6)P₄ experiments was further examined by treating [³H]-I(1,4,5)P₃ with an excess of ATP over the course of an hour and measuring IP₃ consumption and product generation at regular time-points. Even at 300μM IP₃ no IP₅ formation was detected by HPLC for K121W while wild-type enzyme produced both IP₄ and IP₅ within the first ten minutes. Over the course of the experiment the level of IP₄ never exceeded approximately 15% in the wild-type reaction as the level of IP₅ increased. During the K121W reaction IP₄ was also observed within ten minutes, but despite near-complete conversion to IP₄ in the presence of 1mM ATP no IP₅ was observed (Fig 3-4.A).

To identify which IP₄ species was generated larger amounts of [³²P]-IP₄ were produced using the K121W construct and treated with inositol polyphosphate 5-
phosphatase (5-ptase). 5-ptase dephosphorylates I(1,4,5)P$_3$ and I(1,3,4,5)P$_4$, removing the 5'-phosphate, but not I(1,4,5,6)P$_4$, and can be used to distinguish between the two IP$_4$ species. Reactions of K121W Atlpmk with equimolar levels of IP$_3$ and [$^{32}\gamma$]-ATP were run until all ATP was consumed and halted by boiling, then treated with 5-ptase. As an internal control [$^3$H]-I(1,3,4,5)P$_4$ was added to half the reaction before treatment with 5-ptase. After treatment the reactions were separated by HPLC and examined to determine if the K121W-produced [$^{32}$P]-IP$_4$ had been dephosphorylated to IP$_3$. Untreated reactions showed peaks in the [$^{32}$P] and [$^3$H] channels at 15.5 and 17 minutes, confirming presence of the IP$_4$ species and separation at different times indicating two different species. Upon treatment with 5-ptase the [$^3$H]-I(1,3,4,5)P$_4$ peak was reduced and a new peak seen at 13 minutes, indicating that the I(1,3,4,5)P$_4$ had been dephosphorylated to I(1,3,4)P$_3$, as expected and confirming that the 5-ptase was active. Conversely no loss of [$^{32}$P]-IP$_4$ signal was observed after treatment with 5-ptase nor a second peak that could indicate formation of an IP$_3$ species. These tests were repeated with purified [$^{32}$P]-IP$_4$ and the same results were observed (Fig. 3-5.B).
**Figure 3-7: IP₄ production and identification**

*A. Time-courses of IP₄ and IP₅ production.* HPLC traces of [³H]-IP₃ phosphorylation by wild-type or K121W AtIPmk.  
*B. Identity of K121W IP₄ species.* HPLC traces of 5-ptase treated or untreated [³²P]-IP₄ and [³H]-I(1,3,4,5)P₄.

The elution of the K121W AtIPmk [³²P]-IP₄ product at a different time than I(1,3,4,5)P₄ and resistance to dephosphorylation by 5-ptase identifies it as I(1,4,5,6)P₄.

This was the expected product for two reasons. First, AtIPmk has never been observed to generate I(1,3,4,5)P₄ or I(1,2,4,5)P₄ as the product of I(1,4,5)P₃ phosphorylation (Stevenson-Paulik et al., 2002). In fact, no IPMK has ever been reported with a 2-kinase activity, hardly surprising as it would likely require a different active site and binding geometry to accommodate an inositol polyphosphate with the target hydroxyl in the proper orientation. Secondly, I(1,4,5,6)P₄ is the native product of the first step of
I(1,4,5)P₃ phosphorylation by AtIpmk and the K121W mutant was neither designed nor predicted to interfere with that activity, but instead with the ability to use I(1,4,5,6)P₄ as the substrate for the second phosphorylation.

Kinetic parameters of wild-type, K117W, and K121W AtIpmk

While screening indicated that the production of IP₃ by both the K117W and K121W constructs had been inhibited, and as predicted to a greater extent in K121W, the actual degree of change in catalytic efficiency was unknown. To measure the extent of the changes in substrate selectivity the kinetic parameters Kₘ and Vₘₐₓ were measured by steady-state kinetics. Both mutants and wild-type AtIpmk were characterized for the single-phosphorylation reactions upon I(1,4,5)P₃, I(1,3,4,5)P₄, and I(1,4,5,6)P₄. IP₃ experiments used [³H]-I(1,4,5)P₃ while IP₄ experiments used [³²P]-I(1,4,5,6)P₄ generated by K121W AtIpmk or [³²P]-I(1,3,4,5)P₄ by DmIpmk and purified by HPLC.

Compared to the kinetics reported for the GST-fusion form of AtIpmk (Stevenson-Paulik et al., 2002), the wild-type Kₘ for I(1,4,5)P₃ was ten-fold higher at 162µM but also significantly faster at 8.96µmol/min/mg versus 0.052µmol/min/mg. Similar increases were seen for both IP₄ species. Overall catalytic efficiency was higher for all IP species with the new construct. Both mutants exhibited altered kinetic properties for I(1,4,5,6)P₄ and I(1,3,4,5)P₄. For I(1,3,4,5)P₄ velocities were essentially identical between all three forms of AtIpmk, but the Kₘ was 18-fold higher for K117W and 15-fold higher for K121W. Both mutants were also poorly catalyzed the phosphorylation of I(1,4,5,6)P₄ with K121W clearly worse than K117W. Comparing the catalytic efficiencies of each mutant to wild-type the greatest reduction is observed for
the K121W I(1,4,5,6)P$_4$ reaction followed by K117W and the same pattern for the I(1,3,4,5)P$_4$ reactions.

In contrast the phosphorylation of I(1,4,5)P$_3$ by K121W has a slightly lower velocity and $K_M$, resulting in an overall improvement in catalytic efficiency. Surprisingly not only has the K121W mutation reduced the efficiency of the IP$_4$ reactions by over an order of magnitude, it has also slightly improved the IP$_3$ reaction.

![Kinetic curves for AtIpmk constructs](image)

**Figure 3-8: Kinetic curves for AtIpmk constructs**

The kinetic parameters determined for the K117W and K121W constructs provide strong supporting evidence for the inositol binding hypothesis. Both mutants have clearly reduced catalytic efficiencies for I(1,4,5,6)P$_4$ and I(1,3,4,5)P$_4$ and the pattern of losses matches that predicted by the sites blocked and possibility of steric conflict.
relieving side-chain conformations. For K117W the I(1,3,4,5)P$_4$ and I(1,4,5,6)P$_4$ efficiencies are reduced 14.6- and 39.0-fold, while for K121W the reductions are 18.3- and 81.4-fold. K117W was predicted to have similar losses for both IP$_4$ species, with I(1,4,5,6)P$_4$ slightly higher due to the presence of phosphates in both the γ- and δ- pockets, where the δ-phosphate might prevent the γ-phosphate from rotating as freely to relieve any steric clash with K117W. A greater loss of activity was also predicted for K121W versus K117W for I(1,4,5,6)P$_4$ due to the possibility that K117W may be able to adopt a conformation that would relieve the steric conflict with the γ-phosphate. Significantly the loss of efficiency is four-fold different between IP$_4$ species for the K121W construct. Phosphorylation of I(1,3,4,5)P$_4$ was predicted to experience a relatively small degree of loss due to rotamers that could come into conflict with the inositol ring and minor clashes with the γ-phosphate. K121W could be modeled with only a small amount of steric conflict with the γ-phosphate and in the proposed model the δ-pocket would not be occupied by a phosphate. Conversely I(1,4,5,6)P$_4$ was predicted to have phosphates occupying both the γ- and δ-pockets and therefore K121W would be expected to cause a much greater loss of catalytic efficiency. This was exactly the pattern observed.
Table 3-1: Kinetics

<table>
<thead>
<tr>
<th>IP Species</th>
<th>Wild-type</th>
<th>K117W</th>
<th>K121W</th>
</tr>
</thead>
<tbody>
<tr>
<td>I(1,4,5)P₃</td>
<td>( V_{\text{max}} ) (µmol/min/mg)</td>
<td>8.96</td>
<td>NP ²</td>
</tr>
<tr>
<td></td>
<td>( K_M ) (µM)</td>
<td>162.2</td>
<td>50.05</td>
</tr>
<tr>
<td></td>
<td>( K_{\text{cat}}/K_M ) (s⁻¹M⁻¹)</td>
<td>2.97x10⁴</td>
<td>3.78x10⁴</td>
</tr>
<tr>
<td>Change from Wt ¹</td>
<td>NA</td>
<td>0.786</td>
<td></td>
</tr>
<tr>
<td>I(1,3,4,5)P₄</td>
<td>( V_{\text{max}} ) (µmol/min/mg)</td>
<td>6.06</td>
<td>6.61</td>
</tr>
<tr>
<td></td>
<td>( K_M ) (µM)</td>
<td>30.82</td>
<td>490</td>
</tr>
<tr>
<td></td>
<td>( K_{\text{cat}}/K_M ) (s⁻¹M⁻¹)</td>
<td>1.06x10⁵</td>
<td>7.26x10³</td>
</tr>
<tr>
<td>Change from Wt ¹</td>
<td>NA</td>
<td>14.6</td>
<td>18.3</td>
</tr>
<tr>
<td>I(1,4,5,6)P₄</td>
<td>( V_{\text{max}} ) (µmol/min/mg)</td>
<td>11.04</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>( K_M ) (µM)</td>
<td>74.97</td>
<td>254</td>
</tr>
<tr>
<td></td>
<td>( K_{\text{cat}}/K_M ) (s⁻¹M⁻¹)</td>
<td>7.91x10⁴</td>
<td>2.03x10³</td>
</tr>
<tr>
<td>Change from Wt ¹</td>
<td>39.0</td>
<td>81.4</td>
<td></td>
</tr>
</tbody>
</table>

¹ NP – not performed
² Fold difference; \( K_{\text{cat}}/K_M \) (Wt/Mutant)

Of significant surprise is the observation that the efficiency of the K121W IP₃ 6-kinase reaction has actually been slightly improved compared to the wild-type. This was not a predicted outcome. Given the existence of several rotamers observed during the design stage that possessed steric clashes with the inositol ring of any IP species it was predicted that all IP substrates would experience some loss of efficiency. The presence of other rotamers that blocked IP₄ specific phosphates lead to the prediction that the loss of IP₄ activity would be much higher than for IP₃. The source of this improvement is unknown, but it may be speculated that some conformations of K121W in conjunction
with possible main-chain shifts have resulted in packing interactions that improve binding.

These data also do not support the conserved 1’, 4’-pocket motif. \(I(1,4,5,6)P_4\) in the 3-kinase orientation is easily modeled by simply adding a phosphate group to the 6’-hydroxyl of IP\(_3\) while \(I(1,3,4,5)P_4\) can be modeled in two ways. The first flips the inositol ring about an axis bisecting the middle of the 5’-6’ and 2’-3’ carbon bonds. While this preserves the ring geometry, it results in the 6’-hydroxyl being located in the \(\beta\)-pocket and nowhere near the catalytic residues. Rotating 180° orthogonal to the plane of the inositol ring bring the 1’- and 4’-phosphates into the same pockets, but reverses the geometry of the ring. It also results in no phosphate in the \(\beta\)-pocket. More importantly, if \(I(1,3,4,5)P_4\) is modeled in this conformation both the \(\gamma\)- and \(\delta\)-pockets are now occupied. Therefore the K121W mutant would be expected to have similar losses of catalytic efficiency for both species of IP\(_4\), with a similar small difference between IP\(_4\) species as observed for K117W. Most importantly, the 6-kinase phosphorylation of \(I(1,4,5)P_3\) would be predicted to be severely inhibited, as now the 4’-phosphate would come into conflict with K121W in the \(\delta\)-pocket. Clearly this has not occurred.

**Summary**

The identification of the K121W IP\(_4\) product as \(I(1,4,5,6)P_4\), strong preference for IP\(_3\) to IP\(_4\), and kinetic constants all support the inositol binding hypothesis. These data also strongly contradict the proposal for a 1’, 4’-conserved phosphate binding pocket. The pattern of catalytic efficiency loss indicates that the K117W and K121W mutations
are interfering with substrate binding in different manners that are dependent upon the phosphorylation state of the IP species. While the loss of IP$_4$ 6- and 3-kinase activity is not complete, it does suggest that these constructs may be used to probe inositol signaling in vivo or the inositol binding hypothesis used to design additional and improved selective constructs.

**Experimental Methods**

*Design of Substrate Selective AtIpmk Mutants* – To identify residues with the potential to alter the substrate profile of AtIpmk the structures of AtIpmk, ADP-bound ScIpmk (2IF8), and substrate and product bound HsIpk3 (1W2C and 1W2D) were overlaid with LSQMAN (Kleywegt, 1996). The coordinates of I(1,4,5)P$_3$ and AMP-PnP from HsIpk3 were overlaid onto AtIpmk and using PyMOL (DeLano, 2002), IP$_6$ from the structure of ADAR2 (adenosine deaminase acting on RNA) (1ZY7) was overlaid upon the inositol ring orientation of I(1,4,5)P$_3$ to build a library of inositol binding modes. Each mode has a different hydroxyl (6’, 3’, or 5’) occupying the position of the hydroxyl to be phosphorylated. Conserving the ring orientation also conserves the positions of the equatorial hydroxyls and by extension the attached phosphate groups except when replaced by the 2’-axial hydroxyl. Phosphate pockets were defined as the volume of space occupied by an equatorial phosphate group. Residues within 5Å of the putative γ- or δ-phosphate pockets were identified as potential candidates for mutation. K117 and K121 were mutated in silico in KiNG (Chen et al., 2009) to residues of sufficient size and length to occupy the corresponding phosphate pockets. All rotamers were modeled and examined for any steric conflicts (clashes) with the phosphate group in
question. To sample the rotational space around each rotamer in the KiNG rotamer library (Davis et al., 2004), each χ angle was rotated until the “excellent” range had been exceeded or ±10° if the “excellent” range was less than 10°. I(1,4,5)P₃, I(1,3,4,5)P₄, and I(1,4,5,6)P₄ binding in the 6- and 3-kinase orientations were modeled. Mutants were selected by the following criteria: 1. All rotamers without protein-protein clashes must occupy a phosphate pocket. 2. One or two of the three modeled IP species must not have clashes with the pocket-occupying rotamers. Side-chain to side-chain clashes were treated as no-clash rotamers for the purpose of condition one. K117W was predicted to exclude both IP₄ species via the γ-pocket and K121W exclude I(1,4,5,6)P₄ via the δ-pocket. A near-rotamer conformation was observed for K117W that was predicted to partially relieve clash and therefore lessen any selectivity gain. Primers were subsequently designed and protein expressed as previously described.

*Generation of Altered Specificity Constructs* – Wild-type *AtIpmk* in pET-15b was used as the template to perform PCR mutagenesis following the QuikChange protocol. The K117W the primers were: sense, 5’-CT GAA GAA TAC ATC CAA TGG TGT TTG AAG AAA GAC ACG GGT ACC-3’ and antisense, 5’-GGT ACC CGT GTC TTT CTT CAA ACA CCA TTG GAT GTA TTC TTC AG-3’. For K121W the primers were: sense, 5’-C CAA AAA TGT TTG AAG TGG GAC ACG GTG ACC ACA ACC GTG-3’ and antisense, 5’-CAC GGT TGT GTG ACC CGT GTC CCA CTT CAA ACA TTT TTG G-3’.

*Expression and Purification of Altered Specificity Constructs* – Protein was expressed as previously described in chapter 1 for 6xHis-tagged full-length wild-type
*Atlpmk.* Cell harvesting, lysis, and affinity column loading were performed as previously. After washing the protein was batch-eluted with 300mM imidazole and collected in 3ml fractions. Each fraction was spot tested for the presence of protein and positive fractions assayed for size and purity on a 12% SDS-PAGE gel. Fractions with the appropriate sized band (32kDa) were pooled and 1μg of Thrombin per mg eluted protein (measured using the Bio-Rad Protein Assay (#500-0006)) was added to the pooled sample. The sample was transferred into 10kDa SnakeSkin® Pleated Dialysis Tubing (Pierce, #68100) and dialyzed overnight against 2L 50mM Tris-HCl pH7.5, 150mM KCl, 5mM β-Me. After digestion the sample was run over a 500μl p-amino-benzamidine column to remove the thrombin, concentrated to less than 10ml and purified over the Superdex 200 size exclusion column. Elution was tracked by absorbance at 280nm and protein-containing fractions again run on a 12% SDS-PAGE gel. Protein-containing fractions were pooled and aliquoted, flash-frozen with liquid nitrogen, and stored at -80°C.

**Substrate-selective Mutant Product Identification** – The selectivity of the K117W, K121W, and K117W:K121W mutants was screened by phosphorylating I(1,4,5)P₃ in the presence of [³²γ]-ATP and separating the products via Thin-Layer Chromatography (TLC) and HPLC. 10μl reactions of 500μM I(1,4,5)P₃ were treated with 1μg of mutant or wild-type Atlpmk for 30min to 1hr in 50mM Hepes pH6.6, 50mM KCl, 1mM ATP, 10mM MgCl₂, at 30°C and terminated by heat inactivation at 100°C for 1min. 1μl was loaded onto a PEI-F cellulose TLC sheet (VWR # JT4474-4) and ATP, IP₄, and IP₅ species were separated in 1.5M HCl, 1.09M KH₂PO₄, 0.72M K₂HPO₄. The identity of
the K121W IP₄ product was determined by treatment with inositol polyphosphate 5-phosphatase (5-ptase, 5-Pt). Inactivated samples had 10,000 CPM [³H]-I(1,3,4,5)P₄ (American Radiolabeled Chemicals, St. Louis, MO) and 1μl 5-phosphatase or buffer added and were incubated 1hr at 37°C and reactions were halted by the addition of 2μl 2.5M HCl. To test independent of any components remaining from the synthesis reaction 10,000CPM of purified [³²P]-IP₄ and [³H]-I(1,3,4,5)P₄ were added to 10μM cold I(1,4,5,6)P₄ and I(1,3,4,5)P₄ or treated alone and 1 unit 5-ptase or buffer added. Samples were filtered through 0.4μm Acrodisc filters, brought to 500μl 10mM ammonium phosphate pH 3.5 and separated on a 4.6mm X 125mm Partisphere SAX anion exchange column over a 10mM to 1.7M NH₄H₂PO₄ gradient over 25min. Activity was quantitated on [³H] and [³²P] detectors.

Generation and Purification of [³²P]-IP₄ Products – [³²P]-I(1,4,5,6)P₄ and [³²P]-I(1,3,4,5)P₄ were generated as follows. [³²P] labeled IP₄ species were synthesized by phosphorylating 10μM I(1,4,5)P₃ with 10mM ATP and 10μCi [³²γ]-ATP with either K121W AtIPmk to generate I(1,4,5,6)P₄ or Drosophila melanogaster I(1,4,5)P₃ 3-kinase (dmIP3K) to generate I(1,3,4,5)P₄. The reactions were performed in 50mM Hepes pH6.6, 50mM KCl, 10μM ATP, 1mM MgCl₂, at 30°C using 1μg of IPMK or IP3K in a final reaction volume of 10μl and terminated by boiling for 1min. [³²P]-IP₄ products were purified over a 0.2mM HCl to 0.5M HCl, 2mM EDTA, 30-minute gradient on a Partisphere SAX column and 1ml fractions collected. Fractions containing [³²P]-IP₄ were identified and activity quantitated by scintillation counting with ScintiSafe Econo2 (Fisher Chemicals) scintillation fluid in a Beckman LS6500 Multi-Purpose Scintillation
Counter. Samples were neutralized by the addition of 5M KOH until pH7.0 reached and brought to 50mM Hepes pH7.0, aliquoted, and stored at -20°C.

**Inositol Phosphate Kinase Assays** – All unlabeled IPs were purchased from Cell Signals, Inc. Tritiated IPs were purchased from American Radiolabeled Chemicals. Inositol polyphosphate steady-state kinetic reactions were performed in 10μl reactions in 50mM Hepes pH6.6, 50mM KCl, 1.5mM ATP, 5mM MgCl₂, and 0.1mg/ml Bovine Serum Albumin at 30°C and 9.8-1500μM I(1,4,5)P₃, I(1,4,5,6)P₄, or I(1,3,4,5)P₄ with trace amounts of the matching species of [³H]-IP₃ or [³²P]-IP₄. Reactions were halted with 2μl 2.5M HCl, and the products separated by HPLC ([³H]) or TLC ([³²P]). Tritiated reaction traces were quantitated with the Karat32™ program, and [³²P] TLC plates were quantitated with the ImageQuant™ program. All reactions were performed in at least triplicate and data were analyzed with the Graphpad Prism™ program.
Chapter 4. Probing Inositol Signaling *In-vivo*

**Introduction**

One of the principal challenges in understanding inositol signaling has been the difficulty in tracking the roles of different inositol polyphosphate species. This is perhaps most pronounced for those IP species generated by IPMK. While it is not unusual for loss of a gene to result in a lethal phenotype, the difficulty in investigating IPMK related signaling lies in the fact that the IP product or products required for survival cannot be distinguished from each other. Loss of IPMK results in the loss of I(1,4,5,6)P$_4$ or I(1,3,4,5)P$_4$ as well as I(1,3,4,5,6)P$_5$, varying with the organism. While IP3K has been to generate I(1,3,4,5)P$_4$ *in vivo* in *ipmk*Δ yeast (Seeds et al., 2005), the issue remains of I(1,4,5,6)P$_4$ production and the time and location of expression. Previous work has shown that IPMK is expressed in different tissues at different times (Frederick et al., 2005; Nalaskowski et al., 2002) and obviously generation of the wrong IP species in the wrong tissue cannot be expected to illuminate its native role.

Two primary goals existed for this work – determine how IPMK distinguishes between different IP species and develop tools to better investigate inositol signaling *in vivo*. Having redesigned *AtIpmk* with significantly reduced activity against I(1,3,4,5)P$_4$ and I(1,4,5,6)P$_4$, the next stage was to test these designs in a biological setting. Ideally either the K121W construct (K117W deemed not worth testing as, per expectations, it retained greater IP$_4$ activity) would exhibit no IP$_4$ activity *in vivo*, measured by a loss of
IP$_5$ but not IP$_4$ production, or results from *in vivo* studies could be used to further refine a selective design.

Two approaches were selected to measure the efficacy of *At*Ipmk designs in altering the *in vivo* inositol signaling pathways. First, the actual levels of inositol polyphosphates would be measured in *ipmkΔ* yeast strains transformed with plasmids containing various constructs. HPLC analysis of cultures grown in the presence of [$^3$H]-D-­myo-inositol could be used to identify *in vivo* IP pools. By comparing to wild-type, knock-out, and kinase-dead constructs the effects of the different designs on different IP populations could be determined.

Secondly, should there prove to be a difference in inositol production direct phenotype testing could be performed to identify the related species. While yeast may be able to survive without IPMK, they are unable to respond to a number of stress environments. *IpmkΔ* yeast are unable to grow in media containing arginine or ornithine as the sole source of nitrogen, but can be rescued by several different IPMK homologues including plant. Comparison of the rescue of growth by *At*Ipmk constructs to wild-type and negative controls (kinase-dead and empty vector) may be used to identify which IP product is necessary for growth. Previous studies have already shown that I(1,2,3,4,5,6)P$_6$, produced by 2’-phosphorylation of I(1,3,4,5,6)P$_5$ by IPK1, is not a factor in arginine/ornithine utilization. Similarly growth is not rescued by the expression of kinase-dead (D247A) IPMK constructs. These data clearly indicate that a product of IPMK is directly involved in nitrogen metabolism. The same pattern is observed for temperature sensitivity, where *ipmkΔ* yeast cannot grow at 37°C. Therefore production
of either \( I(1,4,5,6)P_4 \) or \( I(1,3,4,5,6)P_5 \) is a necessary step in both cases, but which species for which process is unknown.

**Metabolomics**

Inositol-1,2,3,4,5,6-hexakisphosphate (IP_6) is the principle higher inositol polyphosphate species observed in yeast. Extracts of the soluble fraction of yeast cell lysates grown in the presence of \([^{3}H]\)-inositol show that the primary species by percentage are free inositol and unknown IP_1 and IP_2 species. In ipmk\( \Delta \) lines a large IP_3 peak is observed while peaks for IP_4, IP_5, and IP_6 are not, indicating that with the loss of IPMK the synthesis of IP species from IP_3 is lost. IP_3 along with diacylglycerol (DAG) is principally generated from the cleavage of PIP_2 by Plc-\( \gamma \). Previous work has shown that loss of Plc-\( \gamma \) results in the loss of IP_3 accumulation in ipmk\( \Delta \) cells and in the loss of all higher IP species even when IPMK is present (York et al., 1999). Conversely, an over-production of Plc-\( \gamma \) results in a greater level of IP_6 but no lower IP species, while combination with ipmk\( \Delta \) results in an increased accumulation of IP_3. The phenotypes of plc\( \Delta \) ipmk\( \Delta \) and ipmk\( \Delta \) cells are the same, indicating that accumulation of IP_3 is not responsible for the observed nitrogen metabolic and temperature growth defects.

To test the Atlpmk design efficacy \emph{in vivo} various constructs were cloned into the pRS426 Cup1 vector. This vector contains a +Ura nutritional marker that allows auxotrophic cells to survive on minimal media lacking uracil. Expression of the gene of interest is controlled by the Cup1 copper promoter. Wild-type Atlpmk was cloned into the pRS426 vector using the pET-15b construct as a template and point mutants
generated by PCR mutagenesis. The constructs initially tested were wild-type, K121W, and D247A, the previously reported kinase-dead *Atp* mutant. A c-terminal HA tag was also included. Wild-type and *ipm*Δ yeast cell lines auxotrophic for histidine, leucine, tryptophan, and uracil were transformed and plated onto minimal media lacking uracil and liquid media inoculated from single colonies. After growth in [3H]-inositol labeled media supplemented with copper sulfate the cells were harvested and lysed the soluble fraction analyzed by HPLC to measure the relative levels of inositol polyphosphate species.

As seen previously, the kinase-dead and empty-vector cell lines had no IP species beyond IP3 with an additional accumulation of IP2, suggesting that either IP3 was being generated by a path outside of P1c-γ cleavage of PIP2 or some IP3 was being dephosphorylated to lower IP species (such as by 5- ptase). Expression of wild-type *Atp* rescued IP6 to similar levels as seen for IP3 in the kinase-dead and control lines, and similar to wild-type yeast. K121W expression resulted in production of two higher IP species. Both IP4 and IP6 peaks were observed in the HPLC trace (Fig. 4-1), of similar size to each other and to the IP6 peak seen in wild-type cells. The presence of IP6 was interpreted to indicate that a significant amount of IP5 was still being produced *in vivo*, which was then immediately converted to IP6 by IPK1. This had two immediate implications. First, *in vivo* the K121W construct has a strong preference for IP3 over IP4 as its substrate. If IP4 were the preferred substrate, as observed in the wild-type *in vitro* kinetic data by both the higher velocity and lower K_M, then a significant accumulation, if any, of IP4 would not be expected. Secondly, while the K121W construct is clearly
selective in vivo, it is just as clearly not sufficiently selective to eliminate production of IP₅. The observation of both IP₄ and IP₆ peaks is unique and indicated that there remained potential to develop an IPMK design even more selective for the 6-
’phosphorylation of only IP₃.

Figure 4-1: In vivo inositol polyphosphate populations

Additional examination of the modeled binding of I(1,4,5,6)P₄ with the structure of K121W AtIPmk offers a possible explanation for the retention of some IP₄ 6-kinase activity. During the design process to identify potential selective constructs very restrictive criteria were used in an effort to reduce the computational complexity of the search. One of these restrictions was the decision to not model backbone flexibility, working on the premise that large alterations to the backbone, in addition to being computationally extremely difficult and time consuming, would likely result in a general loss of affinity due to disruption of the inositol binding domain. Returning to the original
design, it was determined that the steric conflicts of the K121W mutation with the $\gamma$- and $\delta$-phosphates could be reduced, though not eliminated, by backbone torsions using the backrub tool in KiNG (Davis et al., 2006). These torsions would not place K121W into steric conflict with the rest of the protein nor move the backbone $\varphi/\psi$ angles out of preferred Ramachandran space. While not able to relive all clash, this modeling does suggest that some local distortion of the IP binding site might be part of the cause for the significant amounts of IP$_6$ observed.

Another, more simple explanation is also possible. As *in vitro* kinetics had already confirmed existence of the IP$_4$ 6-kinase activity, albeit at a greatly reduced rate, it is also possible that the buildup of IP$_4$ *in vivo* simply reach sufficient levels for sufficient time to observe production of IP$_3$ and therefore IP$_6$. In all likelihood a combination of both events is occurring, but it cannot currently be determined with is responsible. More importantly, the continued production of a significant amount of IP$_6$ and therefore IP$_5$ meant that while the inositol hypothesis is further supported by the accumulation of the IP$_3$ product, the K121W construct alone is insufficient to the task of probing inositol signaling.

To address the need for a better construct to probe inositol signaling *in vivo* the K117W construct was reexamined. Alone it clearly has less potential than the K121W construct as it is less specific *in vitro*. Designed to block the $\gamma$-pocket, presumably the existence of a conflict relieving rotamer is responsible for its clear but not complete selectivity for IP$_3$ over IP$_4$. Herein though lays a significant fact – the K117W construct is designed to block the $\gamma$-pocket, while the K121W construct blocks the $\delta$-pocket. These
two designs are intended to block \( \text{IP}_4 \) species by interference at two independent sites. As the goal was to design an \( \text{IP}_3 \) 6-kinase that had no 3-kinase activity toward \( \text{I(1,4,5,6)P}_4 \), what if instead of blocking only one pocket, multiple pockets were blocked simultaneously? The binding model of \( \text{I(1,4,5,6)P}_4 \) predicts that it has phosphates occupying the \( \alpha-, \beta-, \gamma-, \) and \( \delta- \)-pockets. Kinetic evidence strongly supports this model. Therefore a K117W:K121W construct would be expected to much more strongly discriminate against \( \text{I(1,4,5,6)P}_4 \) than either mutation alone. In theory an \( \text{IP}_3 \) 6-kinase, similar to \( \text{IP3K} \), would be the result.

The K117W:K121W \textit{AtIPmk} construct was generated in the pRS426 Cup1 vector and a new round of metabolic labeling experiments carried out. Compared to K121W, the levels of both \( \text{IP}_3 \) and \( \text{IP}_4 \) were increased while \( \text{IP}_6 \) was decreased at least six-fold. Although \( \text{IP}_3 \) production had not been completely eliminated, it was significantly reduced from that seen for the K121W construct design, yet more supporting evidence for the inositol binding hypothesis. In most labeling experiments \( \text{IP}_6 \) production was reduced to levels indistinguishable from the background. To confirm that the different inositol profiles were not due to a lack of expression in the mutants, which were replacing solvent exposed lysine with highly hydrophobic tryptophan residues, western blots were performed for all constructs used in the metabolic labeling (Fig. 4-2). Blotting with \( \alpha-\text{HA} \) for the HA tag at the end of each \textit{AtIPmk} construct and normalizing with \( \alpha-\text{actin} \) revealed practically identical soluble expression for all constructs (wild-type, K121W, K117W:K121W, D247A). This is interpreted to mean that the different inositol profiles
are the result of the changes made to the binding site or kinetic residue and not an artifact of protein expression or stability.

![Western blots of IPMK expression in yeast](image)

**Figure 4-2: Western blots of IPMK expression in yeast**

Although the additional reduction of *in vivo* IP₅ production by the K117W:K121W construct further supports the inositol binding hypothesis, the question remains of how the residual IP₄ 3-kinase activity has been retained. Recombinant protein was generated as was previously described and screened for IP₄ and IP₅ production using IP₃ and [³²γ]-ATP and TLC analysis. Production of IP₄ was significantly reduced from wild-type and both single-mutant constructs and some trace IP₅ observed when reactions were performed with 1mM IP₃ and ATP and 1μg K117W:K121W in a 10μl reaction. IP₅ was not observed in reactions with less substrate or enzyme. A steady-state kinetics curve could not be reasonably determined – activity was lost during incubation at 30°C. This behavior is probably related to the much poorer solubility of the double-mutant, indicating a stability issue. (less than 1mg/ml versus 10mg/ml for K117W and K121W, and over 20mg/ml for wild-type). While clearly the double-mutant is far less soluble *in vitro* than either single mutant or the wild-type constructs, *in vivo* soluble levels are identical. This indicates that in the cellular environment some other factors are
compensating for the reduced solubility conferred by the addition of two highly hydrophobic residues. It may be speculated that as factors exist \textit{in vivo} that improve solubility, there may also exist factors that stabilize conformations of the IP binding site that allow the phosphorylation of I(1,4,5,6)P_4. It is also quite possible that the \textit{in vivo} concentrations of the single- and double-mutant constructs never reaches their respective solubility limits and therefore the reduced solubility seen \textit{in vitro} is of no concern. Regarding interpretation of the inositol profiles, the similar soluble expression levels of \textit{all} constructs indicates that the solubility is not a factor in the different IP₃, IP₄, and IP₆ levels.

\textbf{Phenotypic Investigations}

With two designs with clearly different \textit{in vivo} inositol polyphosphate kinase properties new tools now exist to attempt to probe inositol signaling. While neither is perfect, a comparison of phenotypes between the designed and wild-type constructs could be used to determine if I(1,4,5,6)P₄ or I(1,3,4,5,6)P₅ is the responsible signal. As the IP₄ levels are essentially identical between the K121W and K117W:K121W constructs, rescue of \textit{ipmkΔ} phenotypes regulated by I(1,4,5,6)P₄ would be expected to be the same for each. In contrast, if I(1,3,4,5,6)P₅ is the required species, then a differential rescue would be expected. Wild-type and K121W constructs would be expected to have similar phenotypes, as IP₅ production as measured by the levels of IP₆ is similar for both. However a phenotype closer to \textit{ipmkΔ} would be expected for the K117W:K121W construct, with its significantly reduced IP₆ and therefore IP₅ levels.
One of the most straight-forward measurements of IPMK rescue is survival under selective conditions. When in the presence of arginine or ornithine as the sole source of nitrogen, yeast either grow (wild-type) or do not \((ipmk\Delta)\). If wild-type IPMK is expressed, cell growth as measured by number of colonies observed in serial dilutions is no different than when on a rich media such as YPD. In contrast \(ipmk\Delta\) cells completely fail to grow on nitrogen selective media but like wild-type display no nutrient growth defect on rich media. Previous work has already shown that the rescue of a nitrogen growth deficiency is linked to the kinase properties of several different IPMK species, but the question remains as to which IP species is responsible. Comparison of the growth patterns of the different selective constructs on ornithine-containing media might better identify this species.

Yeast \(ipmk\Delta\) cells were transformed with the pRS426 Cup1 vector containing either wild-type, K121W, K117W:K121W, or D247A \(Atlpmk\) and plated on uracil deficient minimal media plates. A negative expression control cell line, transforming \(ipmk\Delta\) cells with empty vector, was also generated. Individual colonies of each transformant were used to generate ten-fold serial dilutions in sterile water and equal volumes plated onto either minimal-media containing ornithine as the sole source of nitrogen or minimal media with ammonium sulfate and other nitrogen sources (positive control media). To minimize growth variation due to the lower nutritional content of minimal media cells were not grown on YPD plates. After spotting each plate was incubated at 30°C and observed daily for growth.
As expected wild-type cells grew equally well in both media while the empty vector transformant cells grew only on the positive control (YPD) media. The D247A AtIpmk cells also failed to grow on the ornithine media, as expected with no kinase activity and repeating previous observations. The failure of the kinase-dead cells to grow, previously shown by western blot to successfully express IPMK, again supports the previous conclusion that the rescue of ornithine growth is dependent upon the kinase activity of IPMK. Rescue by the K121W AtIpmk construct was complete on the ornithine media but reduced by approximately two-orders of magnitude for the K117W:K121W construct. This growth differential was not observed on the positive-control media plates, indicating that the difference in the kinetic properties of the two construct was responsible (Fig. 4-3).

**Figure 4-3: Nitrogen source growth assay**

Similar experiments were performed testing the rescue of growth of ipmkΔ cells at 37°C. The various AtIpmk construct cell lines were plated onto YPD and uracil-lacking media and grown at 30°C and 37°C. After 2-3 days growth a pattern of rescue similar to the nitrogen-source assay was observed. Wild-type and K121W cells grew equally well at both temperatures while the kinase-dead and vector-control constructs were viable at only 30°C. The K117W:K121W transformed cells had growth at 30°C equivalent to the other constructs, but rescue at 37°C was again reduced approximately two orders of
magnitude (Fig. 4-4). As in the case of nitrogen metabolism, this indicates that the required IP species for survival at 37°C is I(1,3,4,5,6)P₆.

**Figure 4-4: Temperature sensitivity growth assay**

Rescue of growth on ornithine media and at 37°C indicates that I(1,3,4,5,6)P₅ is the IP species required for both phenotypes. The differences in rescue between the K121W and K117W:K121W constructs matches the pattern of IP₄ versus IP₆ levels seen *in vivo*. IP₄ levels in both constructs are similar to those for IP₆ in wild type cells and the K121W construct also has IP₆ levels similar to wild-type. Conversely the amount of IP₆ is significantly reduced in K117W:K121W cells. As IP₆ is the product of phosphorylation of IP₅ by IPK1, it can be inferred that the levels of IP₆ are indicators of the levels of IP₅ being produced by IPMK. If IP₄ was the species responsible for rescue of either phenotype, then it would be expected to see similar growth for both selective constructs on either ornithine media or at 37°C. The observation of a loss of rescue between the two designs indicates that production of IP₅ is the responsible process.

While rescue by the K121W construct is complete for both phenotypes, a question does remain regarding the reduced rescue by the K117W:K121W construct at 37°C. Production and purification of recombinant K117W:K121W *Atlpmk* revealed a 20-fold decrease in solubility versus wild-type, 10-fold versus the K121W construct. At this elevated temperature, is K117W:K121W *Atlpmk* stable? Western blots for the expression of the various constructs were performed with cells grown at 30°C and
showed equivalent expression for all constructs, therefore it can be confidently stated that the difference in rescue of the nitrogen growth phenotype is the result of the different kinase properties of the two constructs. However, what if the loss of growth at 37°C is due to insufficient stability of the K117W:K121W construct? The partial rescue of growth indicates that, at the least, IP_4 is being produced, as the kinase-dead and vector control cells failed to grow at all at 37°C. Unfortunately it cannot distinguish between the possibilities that growth is retarded due to low levels of IP_4 because of a loss of protein stability or due to inherently low levels of IP_3 production.

During kinetic characterization of the wild-type and single-mutant constructs it was determined that activity was steadily lost if the assays were performed at 37°C, but no loss if performed at 30°C. However, this loss affected all three constructs and clearly any loss of activity in the wild-type or K121W protein does not impair rescue at 37°C as growth was equivalent to 30°C. Presumably the cellular environment sufficiently stabilizes these two constructs to allow growth. The concern arises for the K117W:K121W construct due to its greatly reduced overall activity in vitro even at 30°C (Fig. 3-4) and entire order of magnitude lower solubility, even at 4°C, compared to the K121W construct. Liquid media were inoculated with both selective constructs, but surprisingly no growth was observed at 37°C over the course of seven days and therefore neither western blots to measure soluble protein expression nor metabolic labeling experiments to measure IP production could be performed. The reason for a failure to grow in liquid media is unknown, as both clearly grew on solid media distinguished only
by the addition of agar. As such it cannot be definitively stated if rescue of growth at
37°C by the K117W:K121W construct is due to production of I(1,4,5,6)P₄ or
I(1,3,4,5,6)P₅. Efforts to grow both constructs at 37°C in liquid media continue.

Summary

As stated at the beginning of this chapter, investigating the roles of IPMK in
inositol signaling has been complicated by its multi-kinase properties. Simply deleting
IPMK does not allow for distinctions to be made in the roles played by I(1,4,5,6)P₄ or
I(1,3,4,5,6)P₅, or indeed between any of the IPMK products. What has been missing is
an enzyme that catalyzes the phosphorylation of a less broad range of IP species.
Comparisons between these more selective versions of IPMK and the wild-type form
might then allow the identification of roles for different IP species. Following this line
the K121W construct was used to probe in vivo IP synthesis and the initial data used to
design a new, more selective construct, K117W:K121W.

Metabolic labeling confirms that the IPMK constructs are selective for different
inositol polyphosphate species and may be used to probe inositol signaling. Metabolic
labeling with [³H]-inositol reveals that the K121W construct has a reduced ability to
phosphorylate I(1,4,5,6)P₄ as shown by its accumulation. While clearly significant
production of IP₅ still occurs, seen by the production of IP₆ equivalent to wild-type, this
evidence further supports the inositol binding hypothesis and therefore the development
of new constructs based upon it. Following this path the K117W:K121W construct,
designed to block the γ- and δ-pockets, displays similar levels of IP₄ and a significant
reduction in the amount of IP$_6$. While also not a perfectly selective design and clearly less soluble in vitro, when used in conjunction with the K121W construct these two variants of IPMK may be used to probe inositol signaling.

Growth assays on selective media strongly indicate that I(1,3,4,5,6)P$_5$ is the necessary IP species for utilization of ornithine as the sole nitrogen source. Rescue is significantly reduced for $impk\Delta$ cells transformed with the K117W:K121W construct, but complete when transformed with wild-type or K121W. If I(1,4,5,6)P$_4$ was the required species, similar rescue would be expected for all three variants. Western blots of all three as well as the D247A kinase-dead and vector-control transformed cells show that expression of $Atp$pmk is consistent for all forms and therefore the difference in rescue is not due to a lack of expression or solubility. These data further support previous reports that arginine/ornithine metabolism is regulated by inositol polyphosphate species and not by the poly-aspartate loop. Not only does the structure of $Atp$pmk show that it does not contain a similar structural element to the poly-aspartate region, and that it lacks even sufficient residues to form such a loop, but the failure of the D247A kinase-dead construct to rescue also indicates that the kinase functionality is required.

The basis of the rescue of the temperature sensitive phenotype is less clear. While the K121W construct rescues growth at 37°C equivalent to the wild-type, rescue by the K117W:K121W is reduced. This is the same pattern observed for growth on ornithine, but a different concern exists. Western blots and metabolic labeling showed equivalent protein expression and different IP profiles at 30°C, but attempts to grow cells in liquid media at 37°C have failed for both constructs. The reason for failure to grow in liquid
media where the only difference is the lack of agar is unknown. This is a concern specifically for the K117W:K121W construct at 37°C due to the observation that it is significantly less soluble than either wild-type or K121W and therefore raises the question if a lack of protein is responsible for the observed phenotype. Clearly some protein is active given that rescue does still occur, but if rescue is IP₄ dependent the lower degree could be explained by a reduction in IP₄ production due to a reduced level of kinase. Additional changes to the K117W:K121W design may be necessary to improve solubility before it can be used in temperature-sensitivity assays.

**Experimental Methods**

_Yeast Strains and Media_ – Yeast strains (Odom et al., 2000; Seeds et al., 2004; Stevenson-Paulik et al., 2002) were grown in complete minimal media with the appropriate amino acids and 2% sugar. Yeast lines were transformed using the lithium acetate method (Ito et al., 1983).

_Yeast Construct Generation_ – _AtlIpmk_ was cloned into the pRS426 Cup1 vector with a c-terminal HA-affinity tag. Full length _AtlIpmk_ was PCR amplified from the pET-15 construct with primers for SalI and EcoRI N- and C-terminal sites and ligated into pRS426 Cup1 and transformed into XL1-blue cells. The primers were: sense, 5’-CAT GCG GCG GAA TTC ATG CAG CTC AAA GTC CCT GAA C-3’ and antisense, 5’-TAC GTC GAC TCA CTA AGC ATA ATC TGG AAC ATC ATA TGG ATA AGA ATC TGC AGA CTC ATC-3’. Plasmids were sequenced to confirm insertion and no
spontaneous mutations. Point mutants were generated by PCR mutagenesis as previously stated for the pET-15b constructs.

Metabolic labeling – Labeling and analysis were performed as previously described (Stolz et al., 1998). Strains were inoculated from single colonies into 3ml CSM selective media containing 20 μCi/ml [³H]-D-myoinositol (ARC) and 50-150μM CuSO₄ and grown to saturation at 30°C, 250rpm. The cultures were harvested by centrifugation and soluble inositol polyphosphates extracted. Cell pellets were resuspended in 100μl 0.5 M HCl and extracted by vigorous reciprocal shaking in 372μl chloroform:methanol (1:2 v/v) and 100μl glass beads for 2 min. 125μl each of chloroform and 2M KCl were added and the samples shaken an additional 2 min. The organic and aqueous phases were separated by centrifugation at 20,000 X g for 10 min. Soluble inositols in the aqueous phase were resolved by high-performance liquid chromatography (HPLC) on a 4.6mm X 125mm Partisphere SAX anion exchange column, eluted with a 10mM to 1.7M NH₄H₂PO₄ pH3.5 gradient over 25 min. Individual IP isomers were assigned on the basis of co-elution with IP standards.

Western blots – Western blots were performed using mouse anti β-actin (mAbcam 8224) or α-HA (primary) and goat anti-mouse (Odyssey #926-32210) and quantitated on an Odyssey Li-Core thermal imager running software version 2.1. Blots shown are from 12% (vect., wild-type, K121W) or 10% (K117W:K121W, D247A) gels.

Plate growth assays – Single colonies were picked from streaks on selective media, suspended in sterile water, and water-bath sonicated 3x5sec. 5μl of 10-
fold serial dilutions were spotted on CSM-Ura+Dex or Orn-Ura+Dex agar media and incubated at 30°C (control and ornithine) or 37°C (temperature sensitivity).
Conclusions and Future Directions

Conclusions

This work was begun with two primary goals. First, determine the basis of inositol polyphosphate selectivity by Inositol Polyphosphate Multi-Kinase. Secondly, using this knowledge redesign IPMK to create constructs with different IP substrate profiles and use them to probe inositol signaling in vivo.

The structure of Arabidopsis thaliana IPMK has been solved and refined to an R/R\textsubscript{free} value of 23.64/24.61 at 2.9Å resolution. As expected for enzymes with the same activity, the core of AtIpmk is essentially identical to that of ScIpmk and HsIpk3k. No structural component equivalent to the ScIpmk poly-aspartate domain was observed and it is extremely unlikely that AtIpmk can adopt a mimicking conformation. Comparing the sequences of the yeast and plant homologues it is clear that the poly-aspartate domain is part of a larger insert not found in the plant sequence. Residues 231-236 are disordered in the plant structure and the argument might be made that these could adopt a conformation that mimics that of the also disordered yeast poly-aspartate domain. The structure of AtIpmk clearly shows this to be impossible. Residues 230 and 237, leading to and from the disordered region, point in the opposite direction of the residues leading to the disordered poly-aspartate domain. Furthermore these residues and therefore any volume that may be occupied by the disordered loop are on the opposite end of the protein from the disordered region in the yeast structure. Therefore on structural
evidence alone the rescue of growth on ornithine is not performed by an equivalent to the poly-aspartate domain of the yeast homolog. Comparison of the three structures also reveals that the \textit{Sc}Ipmk \(\alpha5\) and \(\beta6\) secondary structure elements are unique and may be of interest regarding the reported protein-complex forming attributes.

As a co-crystal structure of IPMK was not obtainable, it was necessary to use homology modeling to fit \(l(1,4,5)P_3\) from \textit{Hs}Ipm3k into the structure of \textit{At}Ipmk. The ability of IPMK to perform 6-, 3-, and 5-kinase reactions on multiple different IP species indicated that either multiple methods of phosphate addition existed or a common binding motif was used to orient each species in some equivalent manner for catalysis. By comparing all known substrate and non-substrate IP species on the assumption that the position and orientation of the target hydroxyl and ring carbons must be conserved, a binding motif rapidly became apparent. Rather than recognizing IP species by which ring position is phosphorylated, instead it appears that IPMK phosphorylates any substrate that can bind with phosphate groups occupying at least three pockets, two of which are always required and the third is one of two adjacent pockets. Additional selectivity is provided by a strict steric requirement that the ring orientation, and therefore the positions of the target hydroxyl and any phosphate groups, be consistent, and no axial phosphate groups be in steric conflict with the protein.

This motif concept was developed before the structure of \textit{At}Ipmk was solved, but could not be tested until a structure was available that could be used to identify residues that might be mutated to created an IPMK with altered specificities. Simultaneously Chang and Majerus proposed a binding motif for \textit{Hs}Ipmk based on a similar steric
match/exclusion theory. Their motif was slightly different and could not be used to redesign the plant or yeast homologues as it did not accommodate a 6-kinase activity, which is either severely reduced or non-existent in the human homolog.

Various different inositol polyphosphate species were modeled into the active site of \textit{AtIpmk} using the ring and 3’-hydroxyl coordinates of the overlaid IP$_3$ from \textit{HsIp3k} as a scaffold. Each equatorial hydroxyl was modeled in place of the 3’-hydroxyl to generate binding mode models that, by addition or removal of a phosphate group, could be used to generate binding models for any IP species in any putative binding conformation. The lack of 1- and 4-kinase activities may be explained by the observations of these binding modes. In both cases one of the α- or β-phosphate pockets is not occupied. Examination of the \textit{HsIp3k} structure suggests that these phosphate groups may play a role in ordering or activating the network of water molecules and metal ions required for catalysis. These two pockets are also the site of the greatest number of interactions between IP3K and any IP$_3$ phosphate group. Presumably the same holds for IPMK as the residue identity and backbone structure are nearly identical for this region. Therefore any IP species that lacks a phosphate group to fill one of the two pockets is unlikely to bind, or bind in a catalytically active fashion.

Modeling the known IP substrate species it is not surprising that all have phosphate groups occupying the α- and β-pockets and that these are the two phosphate entities conserved in all substrate species according to the proposed motif. As predicted the γ-phosphate group, opposite the kinase target κ-phosphate in the motif, is solvent exposed and has very few possible contacts with the surrounding residues. Likewise the
δ- and ε-phosphate models occupy regions that have some contact with the structure of IPMK, but less than the always required α- and β-phosphates. Multiple steric conflicts between an axial phosphate in the ε-pockets and core of AtIPmk support the rule that any IP species with an axial phosphate in this pocket is prevented from binding. Any IP species with a 2’-phosphate group pointed into IPMK at the δ-pocket would have the κ-hydroxyl inverted and pointed up and away from the orientation seen for IP₃ and presumably not be a target for phosphorylation. Modeling the inositol ring to bring an inverted hydroxyl into the same position as seen for HsIp3k with IP₃ results in either numerous steric conflicts for the phosphate groups or an orientation that has no contacts with IPMK at all. This is further evidence that IPMK must bind the various IP substrates using a common binding motif including ring orientation and therefore that conservation of specific phosphate groups is not a factor in recognition or catalysis.

The structure of AtIPmk does not support the 1’-, 4’-phosphate pocket theory proposed by Ongusaha (Ongusaha et al., 1998; Shears, 2004). If two phosphate binding pockets exist that must be occupied by either the 1’- or 4’-phosphate groups, then at least two different binding modes are required. The 6- and 3-kinase reactions would have the target hydroxyl in the same position relative to the rotation of the inositol ring carbons, but require an inversion of the ring. A second phosphorylation site would be required for the 5-kinase reaction, either on the opposite side of the inositol ring if the ring geometry were maintained, or rotated one ring position if the ring was inverted. Modeling AMP-PnP along with IP₃ into the active site of IPMK clearly indicates that a second phosphorylation pocket would be placed at too far away and at the wrong angle for any
reaction, not to mention that this same site would in the 6-kinase reaction normally be occupied by a phosphate group. It is unlikely that this region would serve as both a phosphate binding pocket for one reaction and the catalytic site for another. The 1’-, 4’-phosphate pocket hypothesis is also not consistent with reports that I(1,4)P₂ and I(1,3,4)P₃ are not substrates. Ironically the 1’- and 4’-phosphates do appear to be required for all IP substrates, but not because they are attached to those specific ring carbons. Rather, to fulfill the phosphate pocket filling needs, these two positions are always phosphorylated but occupy different pockets for different IP species. Hence recognition is not on the basis of which ring carbon has a phosphate group attached, but simply if, when fitting the inositol ring to properly place the target hydroxyl, the required phosphate pockets are occupied.

Steady-state kinetics and metabolic labeling experiments support the inositol binding hypothesis. Examination of the theoretical models of different IP species in the active site of AIIpmk identified the IP binding domain residues Lys117 and Lys121 as candidates for mutation to confer specific altered IP specificity. Mutation of Lys117 to tryptophan was predicted to block the γ-phosphate pocket while Lys121 to tryptophan would block the δ-pocket. Blocking the γ-pocket affinity was predicted to reduce affinity for I(1,3,4,5)P₄ and I(1,4,5,6)P₄ by roughly equivalent amounts as both IP species were predicted to have a (different) phosphate group occupying the same volume as the tryptophan residue. A severe loss of IP₄ affinity was not predicted however due to the existence of a steric conflict relieving conformation in near-rotameric space. Conversely the blockage of the δ-pocket by a K121W mutation was predicted to have a significantly
greater affect on I(1,4,5,6)P$_4$ than I(1,3,4,5)P$_4$. Modeling I(1,4,5,6)P$_4$ into the active site the δ-pocket was occupied by the 1’-phosphate group, but I(1,3,4,5)P$_4$ did not have a phosphate occupying that pocket. Some loss of I(1,3,4,5)P$_4$ activity was still predicted as the δ-blocking conformer also had some minor steric conflicts with the still present γ-phosphate (the 6’-phosphate for this IP species). Both mutants were also predicted to experience a general loss of affinity due to several rotamers that would block binding of any inositol species by interference with the inositol ring itself. Importantly, neither was expected to, aside from this general loss of activity, affect the phosphorylation of I(1,4,5)P$_3$. Neither the γ- nor δ-pockets were occupied when IP$_3$ was modeled for the 6-kinase reaction.

Comparison of the catalytic efficiencies of each mutant for IP$_3$, I(1,3,4,5)P$_4$, and I(1,4,5,6)P$_4$ matched predictions. Similar losses of efficiency were observed for both IP$_4$ species with the K117W construct and I(1,3,4,5)P$_4$ with the K121W construct, while a significantly greater loss was observed for I(1,4,5,6)P$_4$. Intriguingly the efficiency of the 6’-phosphorylation of IP$_3$ by the K121W construct was actually slightly improved over the wild-type. While $V_{\text{max}}$ slightly decreased, there was a proportionally greater decrease in $K_M$ that actually resulted in an overall increase. Without a substrate-bound structure the exact mechanism of this improvement is unknown, but likely packing interactions for the IP$_3$ reaction in the IP binding site have been improved. The pattern of changes in catalytic efficiency also counter the 1’-, 4’-phosphate pocket proposal. If two pockets in the binding site always bind either the 1’- or 4’-phosphate groups, based on the modeling of IP$_3$ from HsIp3k into the active site of AtIpmk both mutants would have had similar
losses of catalytic efficiency for both IP₄ species. More significantly, the IP₃ 6-kinase reactions would also be expected to be severely inhibited, but instead the opposite is observed.

These data are the first direct examinations of inositol substrate selectivity through alteration of the inositol binding domain. No evidence was found that contradicted this inositol polyphosphate selectivity model and therefore the next stage of investigations were begun. With a validated selectivity model and initial selective AtIPmk construct, inositol metabolism and related phenotypes were investigated in S. cerevisiae.

Metabolic labeling experiments of yeast ipmkΔ cells transformed with the various AtIPmk constructs revealed that in vivo production of IP₅ was significantly retarded but not abolished. IP₄ and IP₆ peaks in the K121W cells were equivalent to the IP₆ peak in wild-type cells, indicating that although IP₅ was indeed a poorer substrate than for the wild-type, sufficient activity remained to generate wild-type levels of IP₆. Returning to the design stage the K117W:K121W mutant was constructed to block both the γ- and δ-phosphate binding pockets on the theory that blocking both sites would grant greater specificity than blocking only one. New labeling experiments confirmed this prediction, with greater levels of IP₃ and IP₄ accumulation and significantly reduced IP₆ levels. The cumulative effect of the double-mutant design is another piece of support for the proposed binding hypothesis and also provides a tool for investigating inositol signaling.

While neither selective construct is a perfect IP₃ 6-kinase, the distinct differences in IP₅ production (measured by the downstream IP₆ levels) allows for identification of IP₄
versus IP₅ related phenotypes by differences in rescue of ipmkΔ phenotypes. If the rescue is the same, then I(1,4,5,6)P₄ is the likely species responsible, but if rescue by the K117W:K121W construct is reduced, IP₅ is the likely species. Growth assays performed using ornithine as the sole nitrogen source and at 30°C versus 37°C were performed to determine if these constructs could indeed be used to probe signaling and identify the responsible IP species.

Differential rescue of the ornithine growth phenotype indicates that I(1,3,4,5,6)P₅ is the responsible species, while rescue at 37°C is less certain. While growth at 37°C is fully rescued by the K121W and partially by the K117W:K121W construct, it is unknown if inositol production by the latter is fully functional. Recombinant K117W:K121W AtIpmk solubility is decreased 10-fold and activity in vitro at 30°C could not be quantitated. At 30°C the construct is clearly active in vivo, with equivalent soluble protein levels as measured by western blot to the K121W and wild-type constructs and clear production of IP₄. Liquid cultures at 37°C have failed to grow for unknown reasons, therefore expression of soluble, stable constructs and the levels of IP species are unknown. Until the question of expression and in vivo catalytic activity is answered, it cannot be stated with certainty if IP₄ or IP₅ is the species responsible for survival at 37°C. If both mutants are fully functional in vivo at 37°C, then like rescue of the ornithine growth phenotype, IP₅ is the responsible species.

In summary the primary goals of this work have been achieved. Using the structure of A. thaliana IPMK at 2.9Å and a comparison of the different substrate and
non-substrate inositol polyphosphate species an inositol polyphosphate selectivity hypothesis was proposed. This model was tested by redesigning the inositol binding site and comparing the pattern of changes in catalytic efficiencies to those predicted by the models. The pattern of efficiency loss for different IP₄ species with each mutant matched that predicted. Intriguingly, although a baseline loss of IP₃ 6-kinase efficiency was predicted, a slight overall increase was instead observed for the K121W construct. This is not believed to indicate an error in the model, as the percent change is quite small and can be explained simply by the tryptophan in question easily adopting a binding rather than blocking conformation when both are possible in the presence of IP₃. In contrast the phosphorylation of I(1,4,5,6)P₄ would be expected to always be inhibited as all K121W conformations are predicted to block binding. While the K121W construct did not prove to be sufficiently selective in vivo to probe IP₄ versus IP₅ signaling alone, it is clearly selective against IP₄ and therefore the K117W:K121W construct was designed using the same inositol binding model and exclusion principles. This construct proved to be even more selective, if still not perfectly so, with significantly reduced IP₆ levels. Together these two constructs may be used to probe inositol signaling in vivo and the inositol binding model proposed here to design additional IPMK variants to probe other IP species roles in signaling.

**Future Directions**

Several areas of investigation related to the function of IPMK remain. Foremost of the kinetic properties is the basis of discrimination between the 6’- and 3’-phosphorylation of I(1,4,5)P₃. Comparison of IP₃, I(1,4,5,6)P₄, and I(1,3,4,5)P₄ in the
active site of IPMK has not identified why the plant and yeast forms always
phosphorylate \( I(1,4,5)P_3 \) at the 6’- rather than 3’-hydroxyl. Arg152 may play a role, as in
the plant structure it was observed in the absence of ATP and IP3 to adopt a head-to-head
interaction with the catalytic residue Asp98. However, an equivalent conformation is not
observed for the closest yeast residues, Thr199 or Lys200. The answer may also simply
be that insufficient contacts exists for IP3 to bind in the 3-kinase orientation. \( I(1,4,5,6)P_4 \)
may be able to bind in the same orientation due to the additional contacts made by the 6’-
phosphate group. Mutagenesis studies altering the residues predicted to interact with the
6’-phosphate in the \( \gamma \)-pocket to eliminate potential interactions may help answer this
question. A redesigned binding domain to block the \( \varepsilon \)-pocket and provide additional
contacts in the \( \delta \)-pocket may also help determine the basis of 6’- versus 3’-selectivity.

In regard to using IPMK to investigate inositol signaling \textit{in vivo} improvements are
still needed in the selective IP binding domains. While the K117W:K121W construct has
improved selectivity against IP4 species, some activity clearly remains and there is also a
solubility concern. Additional changes to the IP binding site may improve selectivity if
packing interactions can increase the probability that specific phosphate-binding pockets
are occupied by the K117W or K121W mutations, or potentially other residues. This
would require more complex side-chain conformation calculations and possibly
adjustments of the entire sequence by inserting new residues. A new design process
would be required to address the computational complexity of such an approach.
Solubility might be improved by alterations outside of the active site to compensate for
the penalty of replacing two solvent exposed lysine residues with highly hydrophobic
trypotphan residues. If solubility can be improved presumably the kinetic constants of the new construct could be determined. IP₄ phosphorylation might also be increased as a side-effect, which would require improvements as already described to the IP binding domain.

The greatest area of investigation remaining are the now open possibilities to investigate I(1,4,5,6)P₄ versus I(1,3,4,5,6)P₅ signaling pathways \textit{in vivo}. While some information has been gained from using the K121W and K117W:K121W constructs in yeast, the number of possibilities have not even been scratched for investigating inositol signaling \textit{in vivo}. Using the current constructs or a new generation of designs it should be possible to investigate signaling in multiple organisms including yeast, flies, plants, and mammals. An immediate possibility is the generation of transgenic mice containing the one of the two current designs. Could the embryonic lethal phenotype of \textit{ipmkΔ} mice be rescued by either construct? A new design that is an I(1,3,4,6)P₄ 5-kinase may also be used to investigate pathways dependent upon this IP₄ species.

In conclusion this work has generated novel tools that may greatly expand our knowledge of inositol signaling. In addition to answering the question of substrate selectivity of the multi-kinase, the basis for developing new tools has been tested and proven to contain significant potential. This work was the result of an integrated approach, utilizing techniques across a wide range of fields – protein crystallography, computational design, kinetics, metabolic, and genetics.
Appendix

Figure A-1: Protein precision plus ladder


Figure 1-6


Figures 2-13, 2-14, 3-2, 3-3

Contact dots represent close-contact packing interactions that are not in steric conflict or hydrogen bonds. Contact dots are blue to green as the distance decreases and hydrogen bonds are grey-green. Unfavorable packing distances are represented as spikes, moving from yellow to orange as the distance decreases and the interactions become more unfavorable. Steric conflicts are red spikes and indicate atoms that are too close to be physically possible. Hydrogen atoms were added to the models by the program Reduce (Word et al., 1999) and interactions calculated with Probe (Word et al., 2000). A custom script was used to import the
interaction data into PyMOL for figure generation, actual analysis was performed in the KiNG program.

Figure 3-5


References


Dubois, E., Scherens, B., Vierendeels, F., Ho, M.M., Messenguy, F., and Shears, S.B. (2002). In Saccharomyces cerevisiae, the inositol polyphosphate kinase activity of Kcs1p is required for resistance to salt stress, cell wall integrity, and vacuolar morphogenesis. The Journal of Biological Chemistry 277, 23755-23763.


Jayaraman, T., and Marks, A.R. (1997). T cells deficient in inositol 1,4,5-trisphosphate receptor are resistant to apoptosis. Molecular and Cellular Biology 17, 3005-3012.


Biography

Stuart Tamotsu Endo-Streeter

Date / Place of Birth  November 18, 1978
                        Eugene, OR, USA

Education
    Ph.D Biochemistry, Certificate in Structural Biology and Biophysics, Duke
        University, Durham, NC, October 2009.
        Advisor: John D. York, Ph.D.
    B.S. Biochemistry and Biophysics, Oregon State University, Corvallis, OR, June
        2002.
    B.S. Microbiology, Oregon State University, Corvallis, OR, June 2002.