# A Synthetic Biology Approach for the Characterization of the Inositide Signaling Pathway, and Characterization of an $IP_6$ binding Protein

By

Bradley Palmer Clarke

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry

December 14, 2019

Nashville, Tennessee

Approved:

Charles Sanders Ph.D.
John York Ph.D.
Kathleen Gould Ph.D.
Manuel Ascano Ph.D.
Yi Ren Ph.D.

# DEDICATION

To my loving wife Jane.  Your support and love have been everything to me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

# LIST OF ABBREVIATIONS

Adar2: Adenosine deaminase acting on RNA 2

AP2: Adaptor protein complex 2

AUC: Analytical Ultracentrifugation

BTK: Bruton's Tyrosine Kinase

CCCP: Carbonyl cyanide m-chlorophenyl hydrazone

CD: Circular Dichroism

CDP: Cytidine Diphosphate

Clu1: Clustered mitochondria protein 1

CluA: Clustered mitochondria protein A

CluC: Clu1 Central domain

Cluh: Clu1 mammalian homologue

CluN: Clu1 N-terminal domain

Coi1: Coronatine-insensitive protein 1

DAG: Diacylglycerol

EE: Early Endosome

EM: Electron Microscopy

ER: Endoplasmic Reticulum

Fab1: Forms Aploid and Binucleate cells

$Ga_qQL$: GPCR α subunit constitutively active mutant

GPI: Glycosylphosphatidylinositol

HEAT: Huntingtin, elongation factor 3 (EF3), protein phosphatase 2A (PP2A), and TOR1 repeat

HELCAT: Helical and Catalytic Domain

HPLC: High Performance Liquid Chromatography

IHP: Inositol hexakisphosphate

IolT: Major myo-inositol transporter

$IP_2$: Inositol diphosphate

$IP_3$: Inositol triphosphate

$IP_4$: Inositol tetrakisphosphate

$IP_5$: Inositol pentakisphosphate

$IP_6$: Inositol hexakisphosphate

$IP_7$: Diphosphoinositol pentakisphosphate

Ipk1: Inositol Phosphate kinase 1

Ipk2: Inositol Phosphate kinase 2

IPMK: Inositol Phosphate Multi kinase

LRR: Leucine Rich Repeat

Lsb6: LAS seventeen-binding protein 6

MBP: Maltose Binding Proteins

Mss4: Multicopy Suppressor of Stt4

MVB: Multivesicular Body

NatA: N-terminal Acetylase A

p110: 100kDa catalytic subunit of PI 3 kinase

PAGE: Polyacrylamide Gel Electrophoresis

Pds5: Precocious dissociation of sister chromatids

PI: Phosphatidylinositol

PI(3,4)$P_2$: Phosphatidylinositol 3,4-bisphosphate

PI(3,4,5)$P_3$: Phosphatidylinositol 3,4,5-triphosphate

PI(3,5)$P_2$: Phosphatidylinositol 3,5-bisphosphate

PI(4,5)$P_2$: Phosphatidylinositol 4,5-bisphosphate

PI3P: Phosphatidylinositol 3-phosphate

PI4P: Phosphatidylinositol 4-phosphate

PI5P: Phosphatidylinositol 5-phosphate

Pik1: Phosphatidylinositol kinase 1

PIKFYVE: Phosphatidylinositol Kinase FYVE (Fab1, YOTB, Vac1, Eea1 containing protein)

PIP: Phosphatidylinositol Phosphate

Pis1: Phosphatidylinositol Synthase

PLC: Phospholipase C

PM: Plasma Membrane

PTEN: Phosphatase and tensin homolog

RBD: RNA Binding Domain

Sac1: Suppressor of Actin 1

SE: Sorting Endosome

SEC: Size Exclusion Chromatography

Stt4: STaurosporine and Temperature sensitive 4

SV: Secretory Vesicle

TEV: Tobacco Etch Virus protease

Tir1: Transport Inhibitor Response 1

TLC: Thin Layer Chromatography

TPR: Tetratricopeptide Repeat

UBL: Ubiquitin Like Domain

Vps34: Vacuolar Protein Sorting 34

A: Ala: Alanine

E: Glu: Glutamate

F: Phe: Phenylalanine

G: Gly: Glycine

I: Ile: Isoleucine

L: Leu: Leucine

K: Lys: Lysine

M: Met: Methionine

P: Pro: Proline

S: Ser: Serine

W: Trp: Tryptophan

Y: Tyr: Tyrosine

# Chapter 1 Introduction to Inositide Signaling

## 1.1 The Inositide Pathway

Inositides are a family of signaling molecules largely conserved throughout eukaryotic life, and they play a variety of different roles in a multitude of diverse signaling pathways. While there is great diversity of inositide signaling mechanisms across species, there are fundamental components of the pathway that are highly conserved. The first of these components is the inositol ring. At the center of every inositide is a six-carbon ring called myo inositol. There are other isomers of inositol including scyllo, chiro, muco, neo, allo, epi and cis. Several of these different inositol isoforms can be found in nature but the myo form (seen in Figure 1.1A) is the primary form of inositol found in living organisms and has a unique and interesting structure. The 2-position (highlighted in red in Figure 1.1) on the myo inositol carbon ring has an axial hydroxyl group, while all the other positions are equatorial. With this unique position out of line with the ring, each of the positions on the ring can be uniquely recognized by proteins. However, this interesting feature does not make myo inositol optically active since myo inositol still has a mirror plane of symmetry.

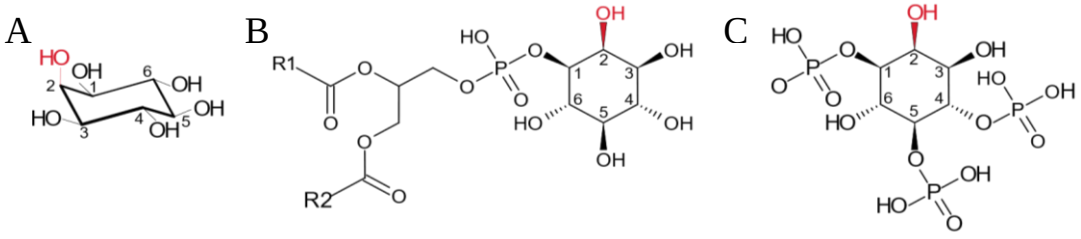


Figure 1.1 Structures of representative inositides. A) Myo inositol represented in chair configuration with axial 2 position represented in red. B) Inositide lipid phosphatidylinositol (PI) where R1 and R2 are commonly often Stearic acid and 20:4(n-6) Arachidonic acid respectively. C) Water soluble Inositide I(1,4,5)$P_3$

It does mean that there are different possible naming schemes for the inositides as D and L Isoforms. Generally, only the D isoform is referred to, and that convention will be adhered to within this work unless otherwise noted. The second important feature of inositides is that each of the hydroxyl positions on the inositol ring can be phosphorylated (and in some cases pyro phosphorylated) to produce a staggering array of possible unique signals. The vast array of possible combinations of phosphates on the inositol ring, and the ability of proteins to specifically interpret these combinations make inositides incredibly dynamic and complex.

One approach to understanding inositide signaling is to manipulate the proteins that produce, modify, interpret and destroy the inositides. In this chapter the different inositides will be discussed and an overview of the various kinases and phosphatases that modify the phosphoinositides will be given, with a focus on the enzymes that will be utilized in later chapters. In Chapter 2, proteins that bind to a particular one of the inositides, the soluble Inositol hexakisphospahte ($IP_6$), will be reviewed. In Chapter 3, development of an expression system for inositide kinases in a clean, noise-less prokaryotic system will be described. Finally, in Chapters 4 and 5, characterization of an $IP_6$ binding protein, Clu1, will be described.

## 1.2 The Lipid Inositides

The inositide signaling pathway has two main branches; the lipid bound phosphoinositide pathway, and the soluble inositol phosphate pathway. The inositol lipids are phosphorylated derivatives of the core lipid phosphatidylinositol (PI) (Figure 1.1B) and are referred to collectively as phosphatidylinositol phosphates (PIPs). The soluble inositides do not have a lipid component and are referred to as inositol phosphates (IPs)

(Figure 1.1C shows I(1,4,5)$P_3$, an archetypal IP). The lipid PIPs are involved in a variety of signaling pathways and play important roles in many of the functions of the endomembrane systems of eukaryotic cells. The PIPs are involved in identifying different cellular compartments, recruiting endocytic machinery in response to signaling events, and responding to extracellular input. A summary of the relative locations and abundances of the PIPs is given in Table 1.1.

**Table 1.1: Phosphatidylinositol Phosphates**

| | Subcellular Location | Relative Abundance |
|---|---|---|
| **PI** | Endoplasmic Reticulum, Plasma Membrane | ~ 80-90% |
| **PI3P** | Early Endosome, Sorting Endosome | ~ 0.05-1.5% |
| **PI4P** | Plasma Membrane, Vesicles, Golgi | ~ 10% |
| **PI5P** | Plasma Membrane, Early Endosome, Golgi | ~ 0.1-0.5% |
| **PI(3,4)$P_2$** | Plasma Membrane, Endocytic Vesicles, Early Endosome | ~ 0.1-1% |
| **PI(3,5)$P_2$** | Sorting Endosome, Multivesicular Bodies | ~ 0.1-1% |
| **PI(4,5)$P_2$** | Plasma Membrane | ~ 10% |
| **$PIP_3$** | Plasma Membrane | ~ 0.1-1% |

Table 1.1 is adapted from (Viaud *et al.*, 2016)

The different PIPs are not present in equal concentrations and are not ubiquitously distributed across the different membrane locations throughout the cell. The distribution of PIPs on different cellular membranes is often thought of as an identification system for the cellular membranes, or as a way for proteins to be recruited to specific membrane compartments by interacting with specific PIPs that are localized to a particular membrane. Generally, one can think of the different PIPs as being distributed across the endomembrane system as a gradient with the least phosphorylated members located more closely to the center of the cell, and the more highly phosphorylated members being located throughout the different membrane compartments and the cell periphery. The least

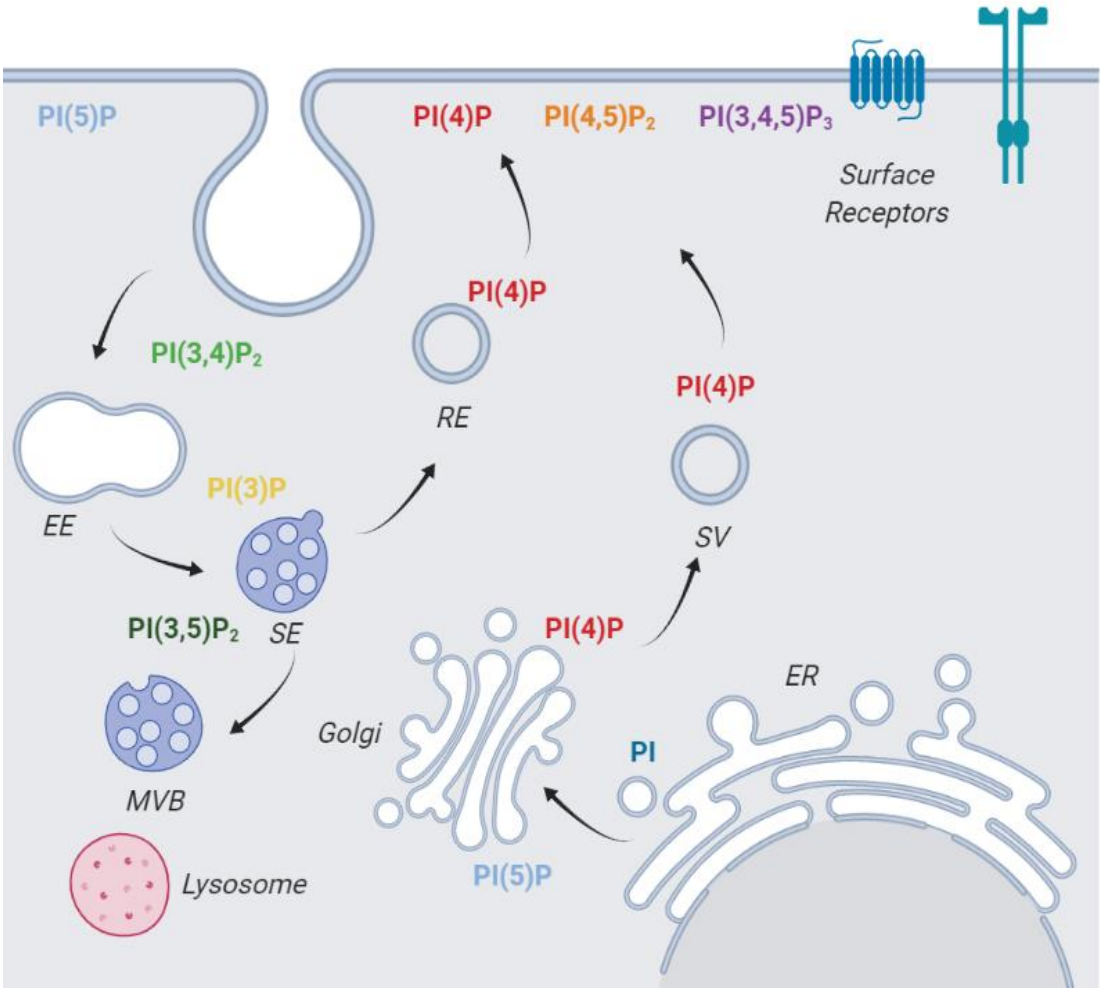


Figure 1.2. Distribution of PIPs in the endomembrane system. ER = Endoplasmic Reticulum, SV = Secretory vesicle, EE = Early Endosome, SE = Sorting Endosome, MVB = Multivesicular body, RE = Recycling Endosome.

phosphorylated of these lipids, PI is primarily synthesized at the endoplasmic reticulum (ER), and ER derived vesicles. Here, near the center of the cell there are few highly phosphorylated PIPs, and this balance is maintained by the presence of phosphatases, which can dephosphorylate a variety of different mono and di phosphorylated PIPs. At the other extreme is the plasma membrane where PI4P is first phosphorylated to $PI(4,5)P_2$, and then further phosphorylated to $PI(3,4,5)P_3$, which in turn can be dephosphorylated to form $PI(3,4)P_2$. PI3P, and the lipids derived from it, $PI(3,5)P_2$, and later PI5P, can be found in

various locations throughout the endomembrane system. However, PIP signaling is highly dynamic, with rapid production and turnover of PIP species coupled with the ever-changing endomembrane system activating and deactivating lipid kinases and phosphatases. While it is helpful to characterize PIPs as marking specific endomembrane compartments, it is also necessary to appreciate the transient nature of these lipids, and how quickly a signaling event can change their distribution. A summary of the relative abundance of each of the phosphoinositides and schematic of their general localizations can be found in Table 1.1, and in Figure 1.2.

### 1.2.1 PI

The most abundant inositol-based lipid is PI (shown in Figure 1.1B). It can represent as much as 10% of total phospholipid content in the cell, and generally is around 80 to 90% of the total inositol lipid. Though PI is synthesized at the ER, specifically on mobile ER associated vesicles, it can be found throughout the other membranes and serve as a precursor to the entire lipid based PIP pathway, and the lipid derived inositol phosphates (Kim, Guzman-Hernandez and Balla, 2011).

### 1.2.2 PI Monophosphates

PI can be phosphorylated at three different positions to produce three distinct phosphatidylinositol mono phosphates PI3P, PI4P and PI5P. The most abundant of these, and first identified is PI4P (Folch, 1942). PI4P represents 10% of the total inositol lipids in the cell and while an important role for PI4P is as a precursor to $PI(4,5)P_2$, it also has signaling roles of its own, and is involved in a variety of Golgi transport pathways (Hama *et al.*, 1999; Walch-Solimena and Novick, 1999). In addition to the plasma membrane,

PI4P is found in many different cellular compartments, and is one of the most widely distributed PIPs. PI3P is found primarily in endosomes and is important for vesicular sorting and is additionally important in regulating cytokinesis. The last identified of the PIPs was PI5P , which represents about 0.1% of total phospholipids (Rameh *et al.,* 1997). It is present at several membrane compartments including the Golgi, early endosomes, as well as the plasma membrane (Sarkes and Rameh, 2010). While its signaling roles are currently less defined than some of the more abundant PIPs, it has been shown to be involved in osmoregulation (Meijer *et al.*, 2001), apoptosis signaling (Emerling *et al.*, 2013) and nuclear signaling (Viaud *et al.*, 2014).

### 1.2.3 PI Diphosphates

PI4P can be further phosphorylated to form $PI(4,5)P_2$, which is primarily located at the plasma membrane. It represents roughly 10% percent of inositol and plays a role in a host of signaling processes. There are many proteins that recognize and bind to $PI(4,5)P_2$, and this binding triggers signaling events involved in endocytosis (Gaidarov and Keen, 1999), regulation of the actin cytoskeleton (Chishti *et al.*, 1998), and exocytosis (Whitaker, 1985; Whitaker and Aitchison, 1985). $PI(4,5)P_2$ serves as the precursor to several very important second messengers, diacylglycerol (DAG) and $I(1,4,5)P_3$, which are released when it is cleaved by Phospholipase C, as well as $PI(3,4,5)P_3$ which is formed from $PI(4,5)P_2$ phosphorylation (Whitman *et al.*, 1988). Conversely, $PI(3,5)P_2$ represents about 0.1-1% of total inositol lipids, and is found throughout endosomes, particularly late endosome where it is important for vesicular sorting (Odorizzi, Babst and Emr, 1998; Dove *et al.*, 2002). $PI(3,4)P_2$ is found at the plasma membrane and plays a role in endocytosis (Zoncu *et al.*, 2007). It is also present in early endosomes, however it is degraded after

endocytosis (Wallroth and Haucke, 2018). It represents about 0.1-1% of total PI lipids.

### 1.2.4 PI Triphosphate

As mentioned above, $PI(4,5)P_2$ can also be further phosphorylated to produce $PI(3,4,5)P_3$, which is an important second messenger involved in cell proliferation (Fruman *et al.*, 1999). $PI(3,4,5)P_3$ is present at almost undetectable levels in quiescent cells, it is rapidly synthesized in response to stimuli and can quickly and transiently increase by 100 fold (Clark *et al.*, 2011). While $PI(3,4,5)P_3$ plays a very important role in cell signaling, it has not been detected in either *S. cerevisiae* or in plants. Despite the lack of $PI(3,4,5)P_3$ in these systems, both yeast and plants have homologues of $PI(3,4,5)P_3$ Phosphatase PTEN (Heymont *et al.*, 2000; Stevenson *et al.*, 2000). So, while it has yet to be observed, it is possible that $PI(3,4,5)P_3$ is present in these organisms, and has simply not yet been detected. Various studies have been aimed at resolving the function of $PI(3,4,5)P_3$, since it plays such an important role in cell growth, and mutations of genes that regulate $PI(3,4,5)P_3$ levels are known to play a role in cancer. In fact, between mutations in PI 3 kinase and $PI(3,4,5)P_3$ phosphatases, misregulation of $PI(3,4,5)P_3$ represents the single most common oncogenic event (Fruman *et al.*, 2017).

### 1.2.5 Other Inositol Lipids

This discussion of inositol lipids focuses primarily on the eukaryotic PIP signaling pathways, however there are other inositol lipids that are separate from the signaling lipids discussed thus far. These other inositol lipids are phosphatidyl inositol mannosides, (the basis of glycosylphosphatidylinositol (GPI)-anchors), as well as inositol ceramides. GPI anchors are a modification utilized as a way to anchor a protein to the membrane.

Ceremide phosphatidyl inositol is found in a variety of eukaryotic species including yeast but, is absent in mammals. While these other inositol lipids are found in eukaryotes, they can also be found in archaea, as well as some bacteria (Michell, 2008). These bacteria and archaea that produce these inositol lipids also produce PI3P and PI, but they are thought to be simply precursors for the other inositol lipids, and not as signaling molecules. Interestingly, the observation that inositol was present in bacterial membranes was made early in the history of inositol lipids with the description of a inositol ceremide lipid in Mycobacterium (Ballou and Lee, 1964).

## 1.3 The Eukaryotic Lipid Inositide Pathway

Like other signaling systems, the inositide signaling pathways consists of, in addition to the signal itself, the cellular machinery to produce, destroy and respond to the signaling molecules. These proteins can be thought of as writers, readers and erasers. The writers are the kinases, that produce the different phosphoinositide signals. The readers are the array of different proteins that recognize and bind to the phosphoinositide signal, and the erasers are the phosphatases that destroy the signal. In the case of the inositide pathway there are also phosphatases that can act as writers, by removing a phosphate to produce an important signal. Additionally, there are also kinases that can act as erasers turning off a less abundant signaling molecule by converting it to a more abundant phosphoinositide. The relative abundance and locations of the different PIPs are maintained by these kinases and phosphatase, and the study of the activities locations and mechanisms of these kinases and phosphatases is central to our understanding of the phosphoinositide signaling pathway. An overview of the pathway is given in Figure 1.3 and Table 1.2.

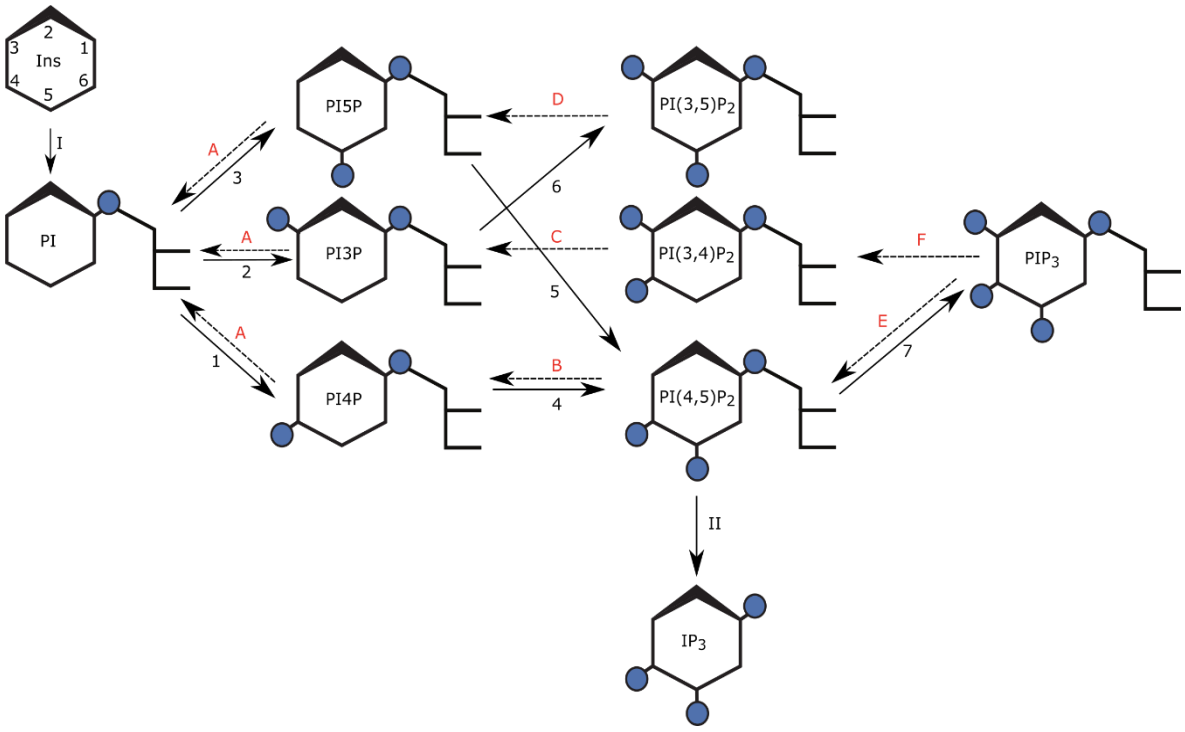


Figure 1.3. Overview of the PIP pathway. All enzymes are described in Table 1.2. Kinases are marked with a number. Phosphatases are marked with red letters, and other enzymes are marked with roman numerals

**Table 1.2: The Enzymes of the Lipid Inositide Pathway**

| **Kinases** | | | **Phosphatases** | | | Other | | |
|---|---|---|---|---|---|---|---|---|
| | *Yeast* | *Mammalian* | | *Yeast* | *Mammalian* | | *Yeast* | *Mammalian* |
| **1** | Pik1 Stt4 Lsb6 | PI4KA/B PI4K2A/B | A | Sac1 | SAC1M1L, SYNJ1/2? | I | Pis1 | PIS |
| **2** | Vps34 | PI3KC3, PI3KC2A | B | Inp51/52/53 | SYNJ1/2 | II | Plc1 | PLCβ PLCγ PLCδ PLCζ PLCη |
| **3** | Fab1 | PIKFYVE | C | | INP4A/B | | | |
| **4** | Mss4 | PIP5K1A/B/C | D | Fig4 | SAC3 | | | |
| **5** | | PIP5K2A/B/C | E | | PTEN | | | |
| **6** | Fab1 | PIKFYVE | F | | SHIP1/2, INPP5E/J/K | | | |
| **7** | | PI3KCA/B/C/G | | | | | | |

Manipulation of the protein regulators of inositide signaling is one of the key methods used to study the functions of inositol signaling. The following section contains a brief review of some of the enzymes responsible for regulation of the inositide signaling pathway, with an emphasis on the proteins studied experimentally in the subsequent chapters. Generally, the names from the yeast homologues of the genes will be used as the majority were first described in yeast, and many of the yeast isoforms are utilized later in this work.

### 1.3.1 Pis1

Reaction I in Figure 1.3/Table 1.2

The attachment of the soluble myo inositol ring to the lipid membrane is a reaction catalyzed by an enzyme called Pis1. The substrates for eukaryotic Pis1 are Cytidine Diphosphate Diacylglycerol (CDP-DAG) and myo-inositol, and the product is phosphatidyl

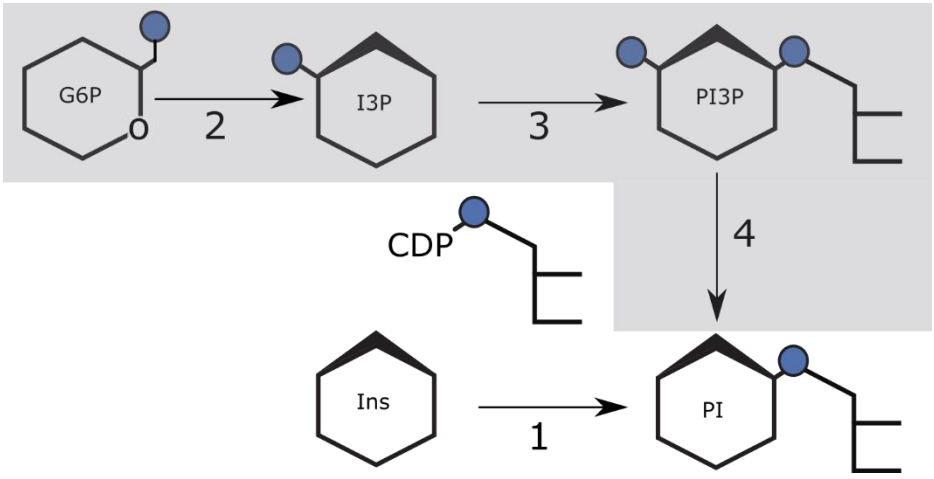


Figure 1.4 Alternative PI synthase pathways in Bacteria/Archaea and Eukaryotes. The Bacteria/Archaea pathway is shaded. CDP-DAG is utilized in both the Bacteria/Archaea and Eukaryotic pathways. Reaction 1, Pis1. Reaction 2, Inositol 1-phosphate synthase. Reaction 3, Phosphatidyl-myo-inositol phosphate synthase (PIP) and Archaeyl-myo-inositol phosphate synthase (AIP). Reaction 3, PIP phosphatase and AIP phosphatase.

inositol (Reaction 1 in Figure 1.4). While most eukaryotes only possess a single isoform of Pis1, plants have two different forms, Pis1 and Pis2 that differ in their preference for CDP-DAG with different acyl chains (Löfke *et al.*, 2008). Pis1 is primarily localized to the ER, more specifically in the ER associated vesicles (Kim, Guzman-Hernandez and Balla, 2011). This ensures that the PI that originates at the ER can be distributed to other locations throughout the cell.

In the various bacteria and archaea that do produce inositide lipids, the synthesis of phosphatidylinositol takes a different route from that found in eukaryotes (the shaded pathway shown in Figure 1.4). In these organisms, I3P is synthesized from Glucose-6-Phosphate by Inositol-1-Synthase (Reaction 2 in Figure 1.4). While this enzyme is called Inositol-1-synthase, this is in reference to the L- isoform rather than the D- isoform (L-1 Inositol Phosphate is the same as D-3 Inositol Phosphate). The I3P then serves as the substrate along with CDP-DAG for bacterial Pis1 (Reaction 3 in Figure 1.4) and PI3P is produced. This PI3P is dephosphorylated by bacterial PIP phosphatase to form PI (Reaction 4 in Figure 1.4) (Morii *et al.*, 2010). There is some indication that the intermediate PI3P may increase under certain circumstances, possibly acting as a signaling molecule (Morita *et al.*, 2010). However, further investigation of this phenomenon will need to be undertaken to determine if this really is an example of prokaryotic PIP signaling rather than just a precursor for other downstream lipid synthesis.

### 1.3.2 The PI Kinases

Since the early days of the study of phosphoinositide signaling, it was known that PI could be converted to PI4P and subsequently to $PI(4,5)P_2$, and that these activities were conducted by two separate enzymes. Therefore, even before their discovery PI 4 kinase

and PI4P 5 kinases were known to exist. PI kinases are an ancient lineage of kinases that emerged in evolution before tyrosine kinases (Manning *et al.*, 2002; Brown and Auger, 2011), and study of how this family of kinases recognize and specifically modify inositide headgroups is of interest both as drug targets and biologically.

PI 4 kinases

Pik1, Stt4, Lsb6, PIKA/B PIK2A/B, Reaction 1 in Figure 1.3/Table 1.2

There are several enzymes that are capable of phosphorylating PI at the 4 position to produce PI4P. The first of these to be cloned and characterized was Pik1 which was cloned from *S. cerevisiae* (Flanagan *et al.*, 1993). It is located primarily at the Golgi and produces the pool of PI4P present within the endomembrane system (Hendricks *et al.*, 1999). The mammalian homologue of Pik1 is known as PI4K2 (the first PI kinase to be identified was later discovered to be a 3 kinase rather than a 4 kinase, thus there is no PI4KI, or rather it was reclassified as a 3 kinase). The next PI 4 kinase to be characterized was Stt4, the kinase responsible for production of PI4P at the plasma membrane (Soshi Yoshida *et al.*, 1994). While deletion of Pik1 from yeast is lethal, stt4 null yeast are still viable, indicating that while the two genes have the same enzymatic activity, they perform two different biological functions, presumably due to differences between their localizations. Lastly, Lsb6 has also been annotated as PI 4 kinase in yeast. It is less well characterized than the other PI 4 kinases and it has less activity than the other members of the PI 4 kinase family. There is also evidence that Lsb6 may serve as a scaffold function for vesicular transport (Han *et al.*, 2002).

PI 3 kinases

Vps34, PI3kC3, PI3KC2A, reaction 2 in Figure 1.3/Table 1.2

In addition to kinase that phosphorylate PI at the 4 position, there are also kinases that add a phosphate to the 3 position. There is a single gene product capable of phosphorylating the inositol ring at the 3 position in *S. cerevisiae*, and that is the PI 3 kinase Vps34 (Schu *et al.*, 1993). Vps34 was identified, and named, because it was discovered through a screen for proteins regulating vacuolar protein sorting. Vps34 is part of a complex that closely regulates the relative levels of PI3P and PI(3,5)$P_2$, and disruption of this complex dramatically impacts a variety of processes that are regulated by the balance of PI3P to PI(3,5)$P_2$ (Stack *et al.*, 1995). While Vps34 has homology to the PI(4,5)$P_2$ 3 kinase catalytic subunit, it has only been shown to have activity against PI, and is not able to further phosphorylate PIPs (Stack and Emr, 1994).

PI 5 Kinase

Fab1, PIKFYVE, reaction 3 in Figure 1.3/Table 1.2

While there is some evidence for the ability of Fab1 and its human homologue PIKFYVE to directly add a phosphate to the 5 position of PI to produce PI5P, their primary activity is to add a phosphate to the 5 position of PI3P producing PI(3,5)$P_2$, which can then be dephosphorylated at the 3 position producing PI5P (Shisheva, 2008). As such they will be discussed more extensively below.

### 1.3.3 The PIP Kinases

As previously mentioned, even before the isolation of PIP kinases, it was known that PI could be sequentially phosphorylated by two separate activities to produce

$PI(4,5)P_2$, thus the discovery of a PI4P 5 kinase was much anticipated. The earliest PIP kinase identified curiously had little activity against PIPs prepared from native membranes but was active against PIP micelles. It was later determined that this was because it was actually PI5P 4 kinase, rather than a PI4P 5 kinase, and the observed production of $PI(4,5)P_2$ was due to contaminations of PI5P in the micelles being 4-Phosphorylated (Rameh *et al.*, 1997). In addition to $PI(4,5)P_2$ producing kinases, PI3P 5 kinases (mentioned briefly above) have also been identified.

PI4P 5 kinases

Mss4, PIP5K2A/B/C, reaction 4 in Figure 1.3/Table 1.2

Because $PI(4,5)P_2$ is such an important signaling molecule it is not surprising that *S. cerevisiae* lacking the only enzyme that converts PI4P to $PI(4,5)P_2$, Mss4, are not viable. Mss4 was first characterized in yeast as a multicopy suppressor of Stt4 (a PI 4 kinase) mutations (Satoshi Yoshida *et al.*, 1994), and later it was identified as a PI4P 5 kinase based on homology (Boronenkov and Anderson, 1995). There are three Mss4 homologues in mammals (α, β, and γ), and while they each have different expression patterns and subcellular localizations only the γ isoform is required for life in mice (Volpicelli-Daley *et al.*, 2010).

PI3P 5 kinase

Fab1, PIKFYVE, reaction 6 in Figure 1.3/Table 1.2

The first PI3P 5 kinase to be characterized is the Yeast protein Fab1 (Gary *et al.*, 1998). This activity was initially discovered when it was observed that osmotically shocked yeast produce elevated $PI(3,5)P_2$, and that this response is dependent of the

function of Fab1 (Bonangelino *et al.*, 2002). Fab1 function is closely regulated by the phosphatase Fig4 (Erdman *et al.*, 1998), and both are part of a signaling complex with the protein Vac14. Together, the members of this complex, along with Vps34, are required for the regulation of the multivesicular body sorting pathway through control of $PI(3,5)P_2$. This Fab1-Fig4-Vac14 complex is conserved in mammals (Dove *et al.*, 2002), and disruptions of the complex by mutations in Fig4 have been identified in patients with both ALS (Chow *et al.*, 2009) and Charcot Marie Tooth disease (Chow *et al.*, 2007).

PI5P 4 kinases

PIP5K2A/B/C, reaction 5 in Figure 1.3/Table 1.2

This family of enzymes is responsible for the conversion of PI5P to $PI(4,5)P_2$. Since the pool of PI5P is much smaller than either PI4P or $PI(4,5)P_2$, these enzymes are thought to be more important for regulating the levels of PI5P than they are for the production of the already very abundant $PI(4,5)P_2$. There are three isoforms (α, β and γ) of PIP5K2, which differ in tissue specific expression, and all are able to form homo, and hereto dimers. One of the isoforms, the γ, has very little activity, and thus it is thought that it may form a hetero dimer with the more active forms (α in particular) acting as a scaffold in order to alter the membrane specific localization of the other isoforms (Clarke, Wang and Irvine, 2010).

### 1.3.3 $PIP_2$ Kinases

PI3KCA/B/C/G, reaction 7 in Figure 1.3/Table 1.2

There is an important family of enzymes in metazoans that phosphorylate $PI(4,5)P_2$ at the 3 position to form $PI(3,4,5)P_3$, and these enzymes are related to the yeast PI 3 kinase

Vps34, though as noted above Vps34 has been shown to only act on PI and not on phosphorylated PIPs (Stack and Emr, 1994). These enzymes consist of a catalytic subunit, p110, and several regulatory subunits. These enzymes are of interest since $PI(3,4,5)P_3$ is an important signaling molecule for a variety of cellular processes including energy processes, survival, and cell cycle progression. There are a variety of cancers that have picked up mutations in this complex, and many of these mutations cluster in the catalytic subunit (Chalhoub and Baker, 2009).

### 1.3.4 Phosphatases

In addition to the kinases responsible for modifying phosphoinositides, there are also many different phosphatases that remove phosphates. Since the PIP phosphatases are very important for the regulation of the inositide pathway a cursory review of some of the important phosphatases and the roles that they play in the phosphoinositide pathway.

#### Sac Phosphatases

One important phosphatase for the regulation of PI mono phosphates is the Sac1 family of phosphatases (A in Figure 1.3/Table 1.2). Sac1 phosphatases can remove a phosphate from any of the monophosphorylated phosphoinositides, as well as removing nonadjacent phosphates from di-phosphoinosidies (i.e. $PI(3,5)P_2$, but not $PI(4,5)P_2$, or $PI(3,4)P_2$) (Guo *et al.*, 1999). This family also includes Fig4 described above which has 3 phosphatase activity against $PI(3,5)P_2$ (reaction D in Figure 1.3/Table 1.2) required for vacuolar sorting (Dove *et al.*, 2002).

5- Phosphatases

In addition to Sac1 phosphatases there are also phosphatases that are specific for phosphoinositides with a phosphate at the 5 position (B in Figure 1.3/Table 1.2). The first of these to be identified was synaptojanin. Synaptojanin is notable for having both a 5-phosphatase domain capable of removing the 5 phosphate from PI(3,4,5)$P_3$, PI(4,5)$P_2$ or PI(3,5)$P_2$, as well as having a Sac domain (Guo *et al.*, 1999). This makes it a quite interesting protein that is capable of first removing the 5 phosphate with one domain, and then further dephosphorylating the remaining position with the Sac1 like domain. Synaptojanin is responsible for regulation of synaptic vesicular processing, and the activity of both domains is required for this function (Mani *et al.*, 2007). Yeast have three proteins that are related to synaptojanin, Inp51, Inp52 and Inp53 (Stolz *et al.*, 1998). Though these enzymes have the same activity, they are localized differently and regulate a variety of cellular processes.

4-Phosphatases

There is also a class of enzymes capable of removing a phosphate from the 4 position (C in Figure 1.3/Table 1.2). In mammals the enzyme that converts PI(3,4)$P_2$ to PI3P is known as Inpp4 and have been shown to act as a tumor suppressor, and is mutated in several types of cancer suppressor (Fedele *et al.*, 2010). It is thought that this is through removal of PI(3,4)$P_2$ and lowering levels of Akt activation (Ivetac *et al.*, 2009). There are also two known bacterial toxins from *Shigella flexneri* and *Salmonella* that act through the removal of the 4 phosphate from PI(4,5)$P_2$, producing PI5P (Niebuhr *et al.*, 2002; Terebiznik *et al.*, 2002).

### 3-Phosphatases

Important for the regulation of $PI(3,4,5)P_3$ are the 3-phosphatases (D in Figure 1.3/Table 1.2), the best studied of which is PTEN. While PTEN is most active against $PI(3,4,5)P_3$, it is also capable of acting on $PI(3,4)P_2$ (McConnachie *et al.*, 2003). By acting to destroy the proliferative factor $PI(3,4,5)P_3$, PTEN acts as an important tumor suppressor (Hollander, Blumenthal and Dennis, 2011). While there are mutations identified in PTEN that are responsible for misvocalization, most of the identified mutations are inactivating mutations to the inositide phosphatase domain (Marsh *et al.*, 1998). Mutations in PTEN have been identified in many types of cancer including glioblastoma, breast cancer, lung cancer, colon cancer, kidney cancer, melanoma, and uterine cancer (Song, Salmena and Pandolfi, 2012; Lee, Chen and Pandolfi, 2018).

### Phospholipase C

In addition to conversion to $PI(3,4,5)P_3$, and degradation by phosphatases, another fate for $PI(4,5)P_2$ is PLC cleavage (II in Figure 1.3/Table 1.2) of the phosphoinositide head group, and release of soluble $I(1,4,5)P_3$ and DAG (Streb *et al.*, 1983). This classical signaling event has served as a textbook example of second messenger signaling. The downstream modification of $IP_3$, and the importance of this will be discussed further below.

## 1.4 Soluble Inositol Phosphate Pathway

The soluble IPs are derived from the lipid bound phosphoinositides when the ester bond between glycerol and the phosphate at the 1 position of the inositol ring is cleaved by a phospholipase C enzyme (Figure 1.5). This results in the formation of the soluble

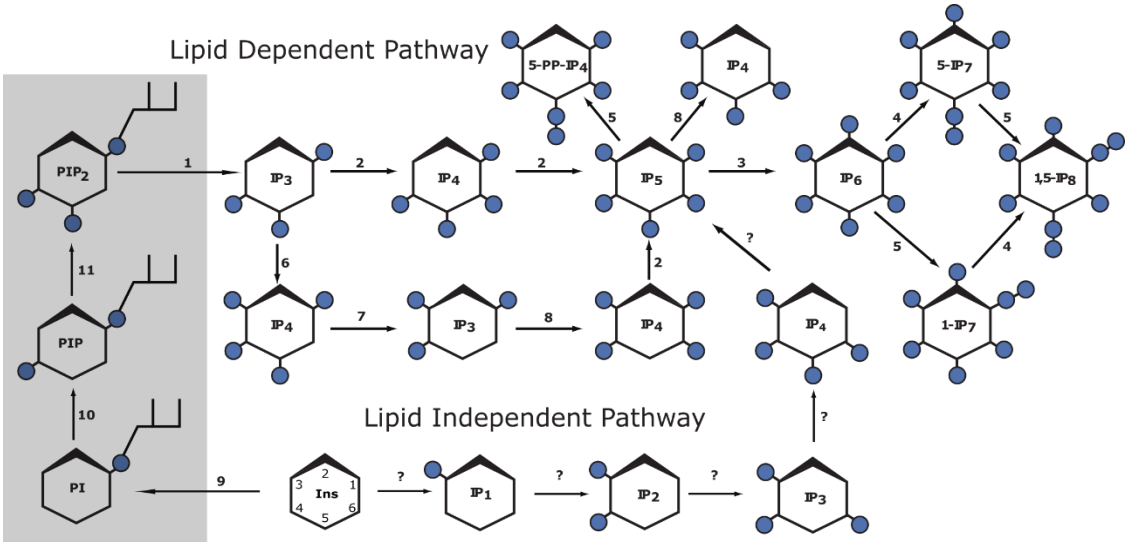


Figure 1.5. The Inositol Phosphate signaling pathway. The upper portion shows the classical Lipid dependent pathway. The lower portion shows the lipid independent pathway found in plants. Enzymes from the pathway are listed in Table 1.3. Figure is based on Hatch and York, 2010.

| **Table 1.3: The Enzymes of the Soluble Inositide Pathway** | | |
|---|---|---|
| Lipid Dependent Pathway | | |
| | *Yeast* | *Mammalian* |
| **1** | Plc1 | PLCβ PLCγ PLCδ PLCζ PLCη |
| **2** | Ipk2 | IPMK |
| **3** | Ipk1 | IPK1 IP5K IPPK |
| **4** | Kcs1 | IP6K1 IP6K2 IP6K3 |
| **5** | Vip1 | Vip1 Vip2 PPIP5K1 PPIP5K2 |
| **6** | | IP3KA IP3KB IP3KC |
| **7** | | Inpp5 |
| **8** | | ITPK IP56K |
| **9** | Pis1 | PIS |
| **10** | Pik1 Stt4 Lsb6 | PI4KA/B PI4K2A/B |
| **11** | Mss4 | PIP5K1A/B/C |

I(1,4,5)$P_3$ being released from the membrane, leaving behind DAG, both of which can act as second messengers. The soluble IP is now no longer associated with the membrane and is able to diffuse throughout the cell and in addition to acting as a signal for calcium released it can also be further phosphorylated. An overview of the soluble IP pathway is found in Figure 1.5, and the key enzymes of the pathway are listed in Table 1.3.

In addition to further phosphorylation this signal can also be terminated by the recycling of I(1,4,5)$P_3$. The I(1,4,5)$P_3$ can either be phosphorylated to produce I(1,3,4,5)$P_4$ and then dephosphorylated to form I(1,3,4)$P_3$ where it can play a role in other signaling pathways, or it can be further dephosphorylated to produce myo inositol that can then be recycled back into phosphoinositide lipids by Pis1 (Storey *et al.*, 1984; Irvine *et al.*, 1986). Additionally, it was also shown that I(1,4,5)$P_3$ can be sequentially phosphorylated by Ipk2 (2 in Figure 1.5/Table 1.3) to form I(1,4,5,6)$P_4$ and then I(1,3,4,5,6)$P_5$ which can also be further phosphorylated by Ipk1 to produce $IP_6$ (3 in Figure 1.5/Table 1.3) (York *et al.*, 1999). $IP_6$ is present at around 50 μM in many different types of cells (Pittet *et al.*, 1989; Barker *et al.*, 2004), and some of the roles of $IP_6$ will be further discussed in Chapter 2.

In addition to kinases that add a phosphate to one of the hydroxyl groups on the inositol ring, there are also kinase that can add a pyrophosphate to certain phosphates as well. Two such enzymes are known to act on $IP_6$, Vip1 (Lee *et al.*, 2007; Mulugu *et al.*, 2007) and Kcs1 (Saiardi *et al.*, 1999) (Reactions 4 and 5 in Figure 1.5/Table 1.3). Vip1 adds a pyrophosphate at the 1 position producing 1-$IP_7$, and Kcs1 adds a pyrophosphate at the 5-position producing 5-$IP_7$. Each enzyme can then act on the $IP_7$ isoform produced by the other enzyme to form 1-5-$IP_8$ (Fridy *et al.*, 2007). There are also enzymes that remove pyrophosphates. Interestingly one of these is the phosphatase domain of Vip1. The dual functional nature of this enzyme capable of both producing 1-$IP_7$ and hydrolyzing 1-$IP_7$ with opposing domains is quite interesting, and future studies of its regulation will be required to determine how these opposing domains are regulated.

The pathway described thus far is the core IP signaling pathway and is conserved throughout eukaryotes (Hatch and York, 2010). In addition to the core pathway there are also other branches of the IP signaling pathway that are present in a subset of eukaryotes.

One of these side pathways is an alternative route to $IP_6$ that is found in metazoans. This pathway relies on the activity of IP3K to add a 3 phosphate to $I(1,4,5)P_3$ to produce $I(1,3,4,5)P_4$ (Reaction 6 in Figure 1.5/Table 1.3). Then the phosphate at the 5 position is removed by Inpp5 producing $I(1,3,4)P_3$ (Reaction 7 in Figure 1.5/Table 1.3). Then Itpk can add a phosphate at the 6 position to produce $I(1,3,4,6)P_4$ (Reaction 8 in Figure 1.5/Table 1.3), which is finally phosphorylated by IPMK to produce $I(1,3,4,5,6)P_5$ (Shears, 2009) (Reaction 2 in Figure 1.5/Table 1.3) which is the substrate for Ipk1, which produces $IP_6$. Another alternative pathway that is observed in plants as well as in slime molds is a lipid independent pathway capable of taking myo inositol all the way to $IP_6$ (the bottom of Figure 1.5/Table 1.3) (Williams, Gillaspy and Perera, 2015). The full nature and enzymes responsible for this pathway are not yet fully characterized.

## 1.6 Summary and Research Aims

Inositides are important signaling molecules that are found in a variety of processes throughout eukaryotic life, and even in select bacteria and archaea species as well. The power behind inositides is the ability of proteins to specifically modify and recognize specific positions around the inositol ring, allowing for a vast array of different signals. The first branch of inositol signaling is based around the lipid PIPs. These lipids are found throughout the membranes of eukaryotic cells, and they play key roles in regulating various vesicular trafficking and a host of signaling events. They are regulated by a large array of kinases and phosphatases that are closely spatially regulated. In addition to the lipids, there are also soluble IPs that are also involved in many different cellular signaling events. The soluble IPs are also regulated by a conserved family of kinases and phosphatases, and there are many emerging roles for them in a variety of different cellular pathways.

The nature of the inositide signaling pathways can make them quite difficult to study. They are so ingrained in life that often times genetic manipulation of the members of the pathway is lethal, and lethal in a variety of ways, making it difficult to determine the exact role they are playing. While some of the enzymes have very specific substrates, and add or remove phosphates from very specific positions, there are other enzymes that are quite promiscuous, and determining the specific activities with *in vitro* assays can at times be difficult, particularly when trying to replicate the reactions of the lipid enzymes with micelle assays. There is also a lot of overlapping activities within the cells. At times there are different enzymes with the same enzymatic activity, but that have different functions due to their different localization. There are problems with both *in vitro* study of these enzymes, as well as *in vivo* study. Many of the key discoveries about this pathway were initially made using the simpler pathway present in yeast. A tool that is an intermediate between the full *in vivo* signaling array found in eukaryotes, and the simple, but unrealistic environment of a micelle assay could prove to be vital to quickly elucidating key elements of phosphoinositide signaling. To this end we have developed a system for expressing both the lipid, and the soluble inositide signaling pathways in the prokaryotic model organism *E. coli*, which possess no native inositide signaling of their own. This system allows for a clean, noiseless system for characterizing the activities of various enzymes in the inositide signaling pathway within a living cell membrane environment. Such a clean system allows for the identification of new activities, and for detection of products that in a living cell might be quickly destroyed by endogenous phosphatases. The development and characterization of this system is presented in Chapter 3.

# Chapter 2 Inositol Hexakisphosphate Binding Proteins

## 2.1 Introduction

Inositol phosphates (IPs) perform many roles within the cell, and they interact with a many different proteins. One of them in particular, $IP_6$, is known to interact with several different classes of proteins. The role that $IP_6$ performs in regulating the structure and function of these proteins can take a variety of forms. There are proteins that modify $IP_6$ by adding and removing phosphates from various positions on the inositol ring. In addition, there are also proteins that primarily interact with the headgroups of phosphatidylinositol phosphates (PIPs) but are also capable of interacting with soluble $IP_6$ as well. There are bacterial proteins, that interact with eukaryotic $IP_6$, exploiting it either as a source of phosphate, or to activate toxins once they enter the $IP_6$ rich environment of the eukaryotic cell. Viral proteins also can adopt a similar strategy, utilizing the host cells endogenous $IP_6$ to assemble and stabilize structures. It has also been shown that there are proteins where binding to $IP_6$ stabilizes them in a particular conformation like hemoglobin, which has been shown to bind $IP_6$ and stabilize the T-state, thus allosterically regulating oxygen affinity. There are also as several proteins that are involved in various stages of mRNA processing, including RNA base editing, splicing and export of mRNA from the nucleus.

The nature of the IP-mediated interactions of these proteins takes a variety of forms, but many of these interactions take place at a positively charged patch on the surface of the protein and are transient in nature. There are also proteins that are bound to highly phosphorylated IPs like $IP_6$ in a much more stable manner, and in binding pockets that are deeper, and less exposed to solvent. These proteins do not seem to be utilizing $IP_6$ as a transient second messenger, but

rather as a cofactor for stability or folding, involved either in the assembly of the protein as a single globular unit, or in the assembly of a larger complex of multiple protein subunits. With its six phosphate groups all concentrated in a small space around the inositol ring $IP_6$ requires a specific pocket with positively charged amino acids that coordinate each of the phosphate groups. There is no consensus motif that proteins have adopted to interact with $IP_6$, in much the same way that $IP_6$ does not have a singular mechanism of action. However, many of the $IP_6$ binding proteins with various functions interact with $IP_6$ utilizing a tandem repeat domain composed of bundles of α helix pairs that create a larger super solenoid, with $IP_6$ binding to a positively charged pocket at the center of the ring. As more proteins are identified as $IP_6$ binding proteins it is likely that the group of proteins utilizing $IP_6$ as a structural cofactor will grow as well, and by studying these interactions we can better understand how $IP_6$ is acting in structural roles within proteins.

This chapter will discuss the nature of these interactions by focusing on the 128 structures of proteins that are annotated as having $IP_6$ (IHP, or I6P) as a ligand on the Protein Data Base (PDB) (Table 2.1). Of these 128 PDBs there are about 35 different proteins represented, and while this list is not comprehensive of all proteins that interact with $IP_6$, it does offer a broad cross section of the kinds of proteins that have been identified and characterized as binding to $IP_6$. The nature of the interactions between $IP_6$ and these proteins based on their structures as well as additional biochemical characterization will be reviewed to better understand the significance of $IP_6$ binding in a variety of biological processes and to provide context for the significance of this work in providing an additional way to study inositide biology.

| Table 2.1: PDBs with $IP_6$ bound | | | |
|---|---|---|---|
| **PDB ID** | **Gene Name** | **Description** | **Citation** |
| ***IP kinases and phosphatases*** | | | |
| **2FVV**, **2Q9P** | Dipp | Diphosphoinositol polyphosphate phosphohydrolase 1 | (Thorsell *et al.*, 2009) |
| **3T9C** | Vip1 | Inositol Pyrophosphate Kinase | (Wang *et al.*, 2012) |
| **2XAL, 2XAM, 2XAR** | Ipk1 | Arabidopsis Inositol 1,3,4,5,6-pentakisphosphate 2-kinase | (González *et al.*, 2010) |
| **3UDZ** | Ipk1 | Inositol pentakisphosphate 2-kinase | (Gosein *et al.*, 2012) |
| **4AQK** | Ipk1 | Inositol 1,3,4,5,6-pentakisphosphate 2-kinase | (Bãos-Sanz *et al.*, 2012) |
| **4LV7** | Ipk1 | Inositol-pentakisphosphate 2-kinase | (Gosein and Miller, 2013) |
| **5MWM** | Ipk1 | Mouse INOSITOL 1,3,4,5,6-PENTAKISPHOSPHATE 2-KINASE | (Franco-Echevarría *et al.*, 2017) |
| **6FJK** | Ipk1 | Arabidopsis Inositol 1,3,4,5,6-pentakisphosphate 2-kinase | (Whitfield *et al.*, 2018) |
| **4O4F** | Ip6k | Inositol hexakisphosphate kinase from *Entamiba histolitica* | (Wang *et al.*, 2014) |
| **6BU0** | PI3KC2α | Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit α | (Chen *et al.*, 2018) |
| ***Lipid interacting proteins*** | | | |
| **1HG5** | AP180 | CLATHRIN ASSEMBLY PROTEIN SHORT FORM | (Ford *et al.*, 2001) |
| **2VGL** | AP2 | ADAPTOR PROTEIN COMPLEX AP-2, ALPHA 2 SUBUNIT | (Collins *et al.*, 2002) |
| **4UQI** | AP2 | AP-2 COMPLEX | (Kelly *et al.*, 2014) |

| | | | |
|---|---|---|---|
| | | SUBUNIT ALPHA-2 | |
| **6QH5** | AP2 | AP-2 complex subunit alpha | (Wrobel *et al.*, 2019) |
| **4AIW** | GPAR1 | GPAR1 | (Van Galen *et al.*, 2012) |
| **1ZSH** | Arrestin 2 | Beta-arrestin 1 | (Milano *et al.*, 2006) |
| **5TV1** | Arrestin 3 | Beta-arrestin-2 | (Chen *et al.*, 2017) |
| **4WPC** | Rgd1p F-BAR | RHO GTPase-activating protein RGD1 | (Moravcevic *et al.*, 2015) |
| **4Y94** | BTK PH-TH | Non-specific protein-tyrosine kinase | (Wang *et al.*, 2015) |
| **2LHA** | Synaptotagmin | Synaptotagmin-1 CB2 domain | (Joung, Mohan and Yu, 2012) |
| ***Bacterial Phosphatases*** | | | |
| **1DKP**, **1DKQ** | Phytase | PHYTASE | (Lim *et al.*, 2000) |
| **3MMJ** | PTP-like phytase | Myo-inositol hexaphosphate phosphohydrolase | (Gruninger *et al.*, 2012) |
| **3NTL** | AgpE | Acid glucose-1-phosphate phosphatase | Not published |
| ***Bacterial Toxins*** | | | |
| **3EEB** | RTX | *V. cholera* RTX toxin RtxA | (Lupardus *et al.*, 2008) |
| **3FZY** | RTX | *V. cholera* RTX toxin RtxA | (Prochazkova *et al.*, 2009) |
| **3GCD** | RTX | *V. cholera* RTX toxin RtxA | (Shen *et al.*, 2009) |
| **3HO6** | TcdA | *C. difficile* Toxin A | (Pruitt *et al.*, 2009) |
| **3PA8** | TcdB | *C. difficile* Toxin B | (Puri *et al.*, 2010) |
| **3PEE** | TcdB | *C. difficile* Toxin B | (Shen *et al.*, 2011) |
| **5KLP**, **5KLQ** | HopZ1a | YopJ effector | (Zhang *et al.*, 2016) |
| **5W3T**, **5W3X**, **5W3Y**, **5W40** | PopP2 | YopJ effector | (Z.-M. Zhang *et al.*, 2017) |
| **6BE0** | AvrA | YopJ effector | (Labriola, Zhou and Nagar, 2018) |
| ***Viral Proteins*** | | | |
| **6BHR**, **6BHS**, **6BHT** | HIVCA | Capsid protein p24, Spacer peptide 1 | (Dick *et al.*, 2018) |
| **6ES8**, **6H09** | HIVCA | Gag protein | (Mallery *et al.*, 2018) |
| **6CUS** | RSV matrix protein | Matrix protein p19 | (Vlach *et al.*, 2018) |
| ***Hemoglobin*** | | | |
| | Hbg | Human | (Arnone and Perutz, |

| | | | |
|---|---|---|---|
| | | deoxyhemoglobin | 1974) |
| **1DKE** | Hgb | HEMOGLOBIN: ALPHA CHAIN | (Bruno *et al.*, 2000) |
| **1GZX** | Hgb | HEMOGLOBIN ALPHA CHAIN | (Paoli *et al.*, 1996) |
| **1NIH** | Hgb | HEMOGLOBIN (NICKELOUS DEOXY) (ALPHA CHAIN) | (Luisi *et al.*, 1990) |
| **1THB** | Hgb | HEMOGLOBIN A (OXY) (ALPHA CHAIN) | (Waller and Liddington, 1990) |
| **1Y8W**, **1YDZ**, **1YE0,** **1YE1**, **1YE2**, **1YEN,** **1YG5**, **1YGF**, **1YH9,** **1YHE**, **1YHR**, **1YIE,** **1YIH** | Hgb | Hemoglobin alpha chain | (Kavanaugh, Rogers and Arnone, 2005) |
| **3HXN** | Hgb | Hemoglobin alpha chain | Not yet published |
| ***RNA processing*** | | | |
| **6BK8** | Spliceosome | U2 snRNA | (Liu *et al.*, 2017) |
| **5Y88** | Spliceosome | Intron-lariat spliceosome | (Wan *et al.*, 2017) |
| **5YZL** | Spliceosome | Post Catalytic Spliceosome | (Wan *et al.*, 2017) |
| **6EXN** | Spliceosome | Post catalytic P complex spliceosome | (Wilkinson *et al.*, 2017) |
| **5XJC** | Spliceosome | Human spliceosome before exon ligation | (X. Zhang *et al.*, 2017) |
| **5MPS, 5MQ0** | Spliceosome | Spliceosome remodeled for ligation | (Fica *et al.*, 2017) |
| **6AH0**, **6AHD** | Spliceosome | Splicing factor 3B subunit 1 | (Zhan *et al.*, 2018b) |
| **6FF4**, **6FF7** | Spliceosome | RNA-binding motif protein, X-linked 2 | (Haselbach *et al.*, 2018) |
| **5Z56, 5Z57, 5Z58** | Spliceosome | Human activated spliceosome | (Zhan *et al.*, 2018b) |
| **5YZG** | Spliceosome | Human Catalytic Step I Spliceosome C complex | (Zhan *et al.*, 2018a) |
| **6QW6**, **6QX9** | Spliceosome | U4 snRNA | (Charenton, Wilkinson and Nagai, 2019) |

| | | | |
|---|---|---|---|
| **6J6G, 6J6H, 6J6N 6J6Q** | Spliceosome | Yeast B-b2 complex | (Wan *et al.*, 2019) |
| **6ICZ, 6ID0, 6ID1** | Spliceosome | Human intron lariat | (Zhang *et al.*, 2019) |
| **6QDV** | Spliceosome | Human post-catalytic P complex | (Fica *et al.*, 2019) |
| **3PEU**, **3PEV**, **3RRM, 3RRN** | Gle1 | ATP-dependent RNA helicase DBP5 | (Montpetit *et al.*, 2011) |
| **6B4H** | Gle1 | Nucleoporin GLE1 | (Lin *et al.*, 2018) |
| **1ZY7** | Adar2 | RNA-specific adenosine deaminase B1, isoform DRADA2a | (Macbeth, Schubert, VanDemark, Lingam, Hill, Bass, *et al.*, 2005) |
| **5ED1**, **5ED2**, **5HP2, 5HP3** | Adar2 | Double-stranded RNA-specific editase 1 | (Matthews *et al.*, 2016) |
| **6D06** | Adar2 | Double-stranded RNA-specific editase 1 | (Monteleone *et al.*, 2019) |
| ***Plant hormone receptors*** | | | |
| **2P1M**, **2P1N**, **2P1O, 2P1P**, **2P1Q** | Tir1 | SKP1-like protein 1A | (Tan *et al.*, 2007) |
| **3C6N**, **3C6O**, **3C6P** | Tir1 | SKP1-like protein 1A | (Hayashi *et al.*, 2008) |
| ***N-Terminal Acetylation*** | | | |
| **4HNW** | NatA | N-terminal acetyltransferase A complex subunit NAT1 | (Neubauer, 2012) |
| **4XNH** | NatE | Yeast N-terminal acetyltransferase NatE ($IP_6$) in complex with a bisubstrate | Not yet published |
| **6C95**, **6C9M** | NatA | N-alpha-acetyltransferase 15, NatA auxiliary subunit | (Gottlieb and Marmorstein, 2018) |
| **6O07** | NatE | Naa15 | (Deng *et al.*, 2019) |
| ***Chromatid Cohesion*** | | | |
| **5HDT** | Pds5 | Sister chromatid cohesion protein PDS5 homolog B | (Ouyang *et al.*, 2016a) |
| ***Other*** | | | |
| **1BQ3** | Pgm1 | PROTEIN (PHOSPHOGLYCERATE MUTASE 1) | (Rigden *et al.*, 1999) |

| | | | |
|---|---|---|---|
| **3W8L** | CK2α | Casein kinase II subunit α | (Lee *et al.*, 2013) |
| **5CHV** | CK2α | Casein kinase II subunit α | (Brear *et al.*, 2016) |
| **5MOE**, **5MPJ** | CK2α | Casein kinase II subunit α | (De Fusco *et al.*, 2017) |
| **5OQU, 5ORJ** | CK2α | Casein kinase II subunit α | (Iegre *et al.*, 2018) |
| **5ICN** | HDAC | Histone Deacetylase MTA1 complex | (Watson *et al.*, 2016) |
| **5IJJ, 5IJP** | Vtc4 | SPX domain of C. thermophilum Vtc4 | (Wild *et al.*, 2016) |
| **5MZA** | PfEMP1, ICAM-1 | Erythrocyte membrane protein 1 (PfEMP1) | (Lennartz *et al.*, 2017) |
| **6A73** | Cop9 | COP9 signalosome complex subunit 2, Endolysin | not yet published |
| **2K8R** | aFGF | Heparin-binding growth factor 1 | not yet published |
| **6REY** | Proteasome | Human 20S-PA200 Proteasome complex | (Toste Rêgo and da Fonseca, 2019) |

## 2.2 Inositol Kinases and Phosphatases

One major class of proteins that bind to $IP_6$ are the kinases and phosphatases that modify inositides. Because of the nature of the IP signaling pathway, the kinases and phosphatases involved must be able to distinguish specific phosphorylation patterns and add phosphates to the correct position with a degree of specificity. Ipk1 is an enzyme that must recognize $I(1,3,4,5,6)P_5$, and specifically add the final remaining phosphate to the 2 position to form $IP_6$ (York *et al.*, 1999). In order to do this Ipk1 forms interactions with the 4-, 5-, and 6-, positions for initial recognition of the substrate, with additional contacts with the 1-position and 3-position stabilizing interaction with the substrate (Gosein *et al.*, 2012). In a similar fashion the $IP_6$ kinase Vip1 forms extensive interactions to recognize its substrates $IP_6$ and 5-$IP_7$ (Wang *et al.*, 2012). The enzyme that hydrolyzes the pyro-phosphorylated inositides, Dipp, uses a Nudix fold that is conserved in a variety of diphosphate phosphatases to specifically recognize the substrates 1-$IP_7$,

5-$IP_7$, and $IP_8$ (Thorsell *et al.*, 2009). Not all inositide kinases are highly specific, with some able to act as mulitkinases binding to a variety of substrates. One such enzyme is the Ipk6 enzyme from *Entamoeba histolytica* which is capable of phosphorylating both $IP_6$ as well as $I(1,4,5)P_3$. In order to achieve this, it has a binding pocket that binds to $IP_6$ oriented at one angle, but to $IP_3$ with the substrate tilted 55° relative to the binding position of $IP_6$ (Wang *et al.*, 2014). This ability of $IP_6$ binding proteins to offer both a degree of specificity as well as having the ability to recognize different substrates demonstrates how different mechanisms of recognizing $IP_6$ can provide variation to these signaling pathways.

## 2.3 Lipid Binding Proteins

Another category of proteins that bind $IP_6$ are lipid binding proteins. These proteins bind to the Inositol headgroup of phosphoinositides, often $PI(4,5)P_2$, and play roles in signaling events or endocytosis. Some of these proteins are also capable of binding to and recognizing soluble inositol phosphates as well. While some of these interactions may simply be due to the similar shape and orientations of the phosphate groups and have not yet been demonstrated to have any functional significance, there are other cases where soluble $IP_6$ binding has been demonstrated to be significant and even occupy a separate binding site from the PIP binding site. Arrestin 2 (Milano *et al.*, 2006) and Arrestin 3 (Chen *et al.*, 2017) which regulate the activity of G-Coupled Protein Receptors, have been shown to bind to $IP_6$, and furthermore that $IP_6$ binding induces oligomerization and activation of the proteins.

$IP_6$ has also been shown to bind to, and provide thermal stability to the CB2 domain of Synaptogamin, a synaptic vesicle membrane protein (Joung, Mohan and Yu, 2012). Interestingly the binding site for $IP_6$ in Synaptogamin partially overlaps with its binding site for the clathrin

adaptor protein AP2 (Joung, Mohan and Yu, 2012), and this interaction may eventually reveal a role for further $IP_6$ regulation of clathrin mediated endocytosis. Bruton's tyrosine kinase (Btk), a protein that is require for the function of B cells, is activated when it is recruited to membranes through its binding to $PIP_3$. However, it has also been shown that Btk can bind to $IP_6$ (Figure 2.2.1), at a site that is different from its $PIP_3$ binding site, and that binding of $IP_6$ induces a temporary dimerization that activates the protein (Wang *et al.*, 2015).

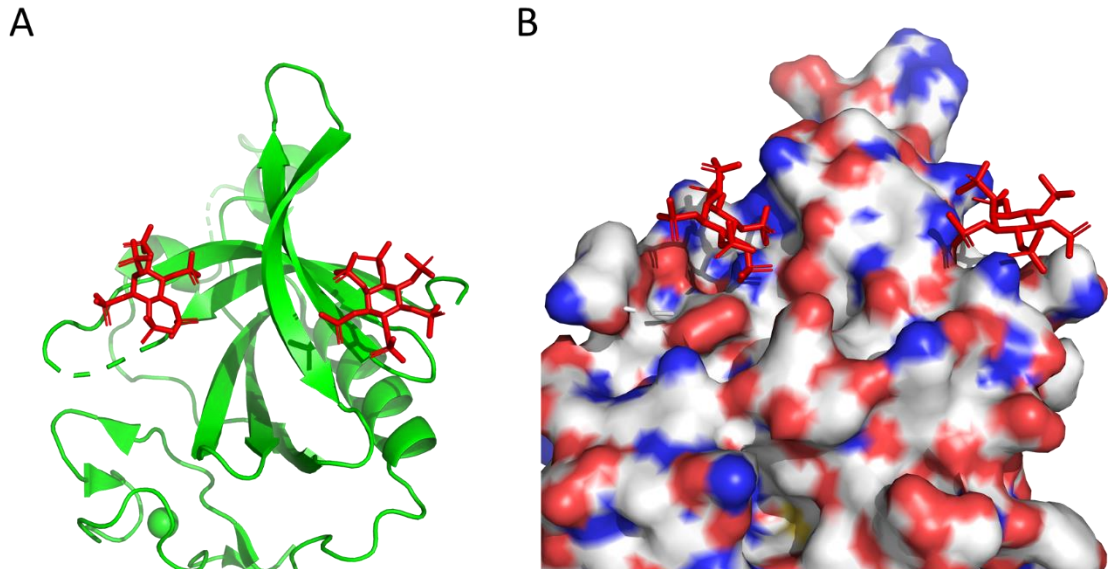


Figure 2.1 $IP_6$ bound to BTK (PDB 4Y94). A) Surface representation showing 2 molecules of $IP_6$ bound to BTK. B) Surface representation of $IP_6$ bound to positively charged patches on the surface of BTK

## 2.4 Bacterial Phosphatases

In addition to phosphatases that modify phosphoinositides as part of eukaryotic inositide signaling, another group of phosphatases that bind to $IP_6$ are the bacterial phytases. Due to the high concentration of phosphates around the inositol ring, $IP_6$ and other highly phosphorylated inositides can serve as a rich storage of phosphates, particularly in plants such as cereal crops

where the phosphate in the form of $IP_6$ in seeds typically accounts for around 2% of their weight, and around 80% of total phosphate content (Reddy, Sathe and Salunkhe, 1982). While most prokaryotes do not have enzymes for IP or PIP signaling, $IP_6$ is an attractive potential source of phosphates for bacteria to utilize as well, and as such bacteria have developed enzymes to exploit this source. Bacterial phytase AppA is a periplasmicly secreted acid phosphatase from *E. coli* that has very high specific activity for hydrolyzing $IP_6$ (Lim et al., 2000). With $IP_6$ as such a rich source of phosphates, bacteria have evolved several classes of phytase including the Protein-tyrosine Phosphatase like phosphatase PhyA as well (Gruninger et al., 2012).

## 2.5 Bacterial Toxins

In addition to utilizing $IP_6$ as a phosphate source, certain bacteria have also developed ways to utilize $IP_6$ as an activator of toxins. Often times when bacteria produce a toxin, they first produce an inactive precursor form of the toxin, that will only be activated later to prevent damage to the bacteria (Gordon and Leppla, 1994). One strategy that bacteria have adopted to promote activation of the toxin only once it is within a eukaryotic cell is to utilize $IP_6$, not found in bacteria but found in abundance within eukaryotes, as an activation factor for toxins. Two such toxins that bind to and are activated by $IP_6$ are the *V. cholera* RTX protease (Lupardus *et al.*, 2008), and the *C. difficil* TcdA/B protease (Pruitt *et al.*, 2009). Both of these toxins are produced as inactive precursors, and upon binding to $IP_6$, the protein is stabilized in a conformation that allows for autocatalytic processing to produce the active form of the toxin (Lupardus *et al.*, 2008; Prochazkova *et al.*, 2009; Pruitt *et al.*, 2009; Shen *et al.*, 2011).

Another class of bacterial pathogen that utilizes $IP_6$ for activation is the YopJ family of acetyltransferases. These acetyltransferases add an acetyl group to host proteins, and this

modification leads to misregulation and impaired function. This family of bacterial proteins do not have significant homology to mammalian acetyltransferases. Both plant pathogens like *P. syringae* and its HopZ (Zhang *et al.*, 2016), and mammalian pathogens like *Salmonella typhimurium* and its AvrA (Labriola, Zhou and Nagar, 2018) YopJ homologues have been shown to bind to $IP_6$ at positively charged patches on their surface. Binding of $IP_6$ allosterically activates the protein by inducing a conformational shift that promotes biding of its required substrate AcCoA, thus increasing its activity (Zhang *et al.*, 2016).

## 2.6 Viral Proteins

Similar to the way bacterial toxins can utilize host $IP_6$ as an activator, viral proteins have also been identified that utilize $IP_6$. It was observed that the HIV capsid contained positively charged pores on the surface, and that these pores were required for infection (Jacques *et al.*, 2016). Later, structures of the HIV Capsid hexamer were solved that revealed $IP_6$ bound at the

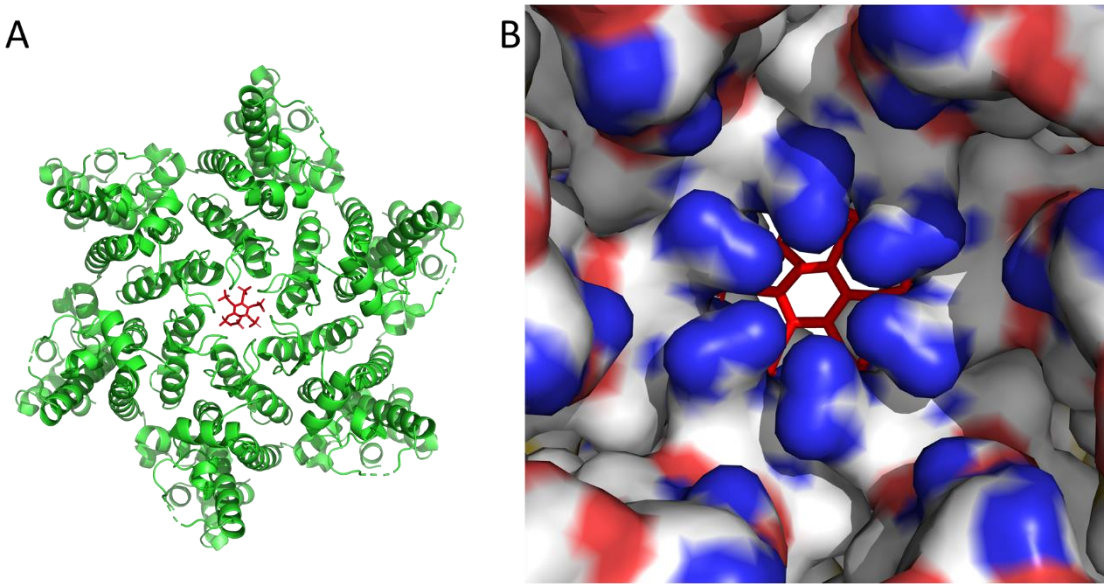


Figure 2.2 $IP_6$ bound to the HIV capsid hexamer (PDB 6ES8). A) Ribbon representation of the HIV hexamer showing that $IP_6$ is bound in the center. B) Surface representation of the positively charged $IP_6$ binding pocket.

center of these positively charged pores (Figure 2.2), and further that $IP_6$ binding to the capsid dramatically stabilized the hexamer and promoted accumulation of viral DNA within the capsid. The $IP_6$ bound form of the capsid hexamer is much more thermally stable, and has a much longer half-life than the apo form (Mallery *et al.*, 2018). Further studies showed that in addition to promoting stability, $IP_6$ promotes assembly of capsid hexamers, and maturation of the HIV capsid lattice (Dick *et al.*, 2018).

## 2.7 Hemoglobin

The earliest structure of a protein bound to $IP_6$ was reported in 1974 by Arnone and Perutz (Arnone and Perutz, 1974). Their structure showed $IP_6$ bound to the center of deoxyhemoglobin (Figure 2.3A and B) stabilizing the T state and decreasing the affinity for oxygen. In many of the subsequent structures of hemoglobin, the $IP_6$ density has been quite weak (Waller and Liddington, 1990; Kavanaugh, Rogers and Arnone, 2005), however the roles of $IP_6$ and $I(1,3,4,5,6)P_5$ as allosteric regulators of hemoglobin that stabilize the T state have been studied in a variety of species (Coates, 1975).

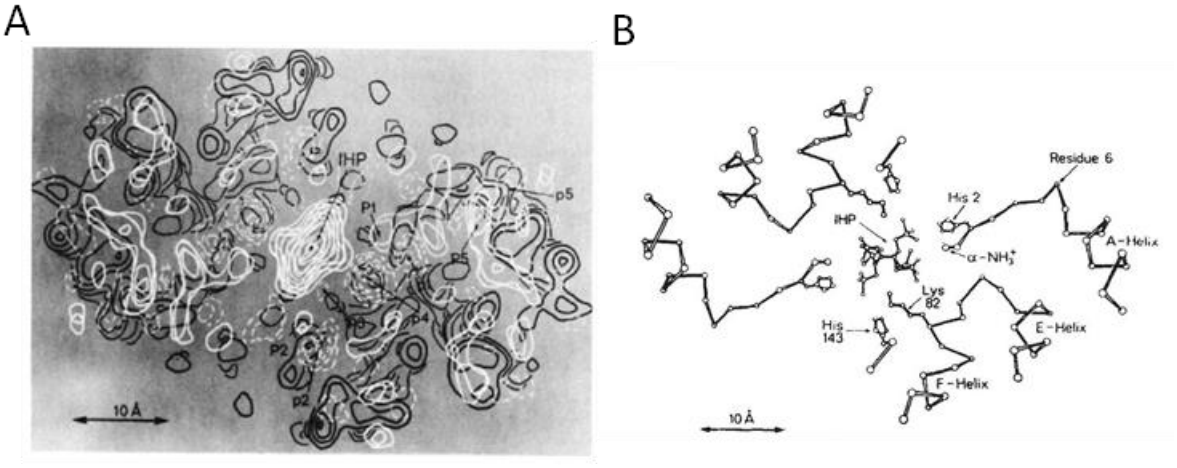


Figure 2.3. $IP_6$ (labeled IHP) bound to Hemoglobin. From Arnone & Perutz 1974. A) Fourier Map and B) model of $IP_6$ binding site. Copyright 1974 Springer Nature

## 2.8 RNA Processing

$IP_6$ has been shown to bind to several proteins that play a role in mRNA processing. The first of these is Adar2 (Macbeth, Schubert, VanDemark, Lingam, Hill, Bass, *et al.*, 2005). Adar2 is a deaminase that converts adenosine to inosine (A to I editing), and in doing so makes a new splicing site that can lead to alternative splicing in some of its substrates (Solomon *et al.*, 2013). The role that $IP_6$ plays in Adar2 is discussed further below. Another RNA processing step that $IP_6$ has been implicated in is RNA splicing. In both the human (Zhan *et al.*, 2018b) and *S. cerevisiae* (Wan *et al.*, 2019) spliceosomes $IP_6$ has been identified as a cofactor. In both cases it binds to Prp8, and in the case of the human spliceosome it appears to stabilize the interaction between Prp8 and the U5 snRNA (Zhan *et al.*, 2018b). After splicing the RNA must exit the nucleus, $IP_6$ has also been demonstrated to play a role in this process by regulating the Gle1 nuclear export protein (York *et al.*, 1999). $IP_6$ binds to Gle1 and Dbp5 near their interface, and the binding of Gle1 to $IP_6$ promotes the export of RNA from the nucleus by stimulating the ATPase activity of Dpb5 (Montpetit *et al.*, 2011).

## 2.9 Structural Binding of $IP_6$

Inositol phosphates are classically studied as signaling molecules, and for many of the IPs this characterization as classical second messengers is a good description. For $IP_6$, some of its interactions may fit into the model of transient biding and serving as a signaling molecule. However, $IP_6$ itself does not always seem to behave in the way that a classical second messenger would be expected to behave. In order for robust signaling to be achieved one would want a molecule that has very low levels at resting state, and then upon stimulus dramatically increases, then after the signal is gone drops back down to low levels. As seen in Figure 2.4 however, this

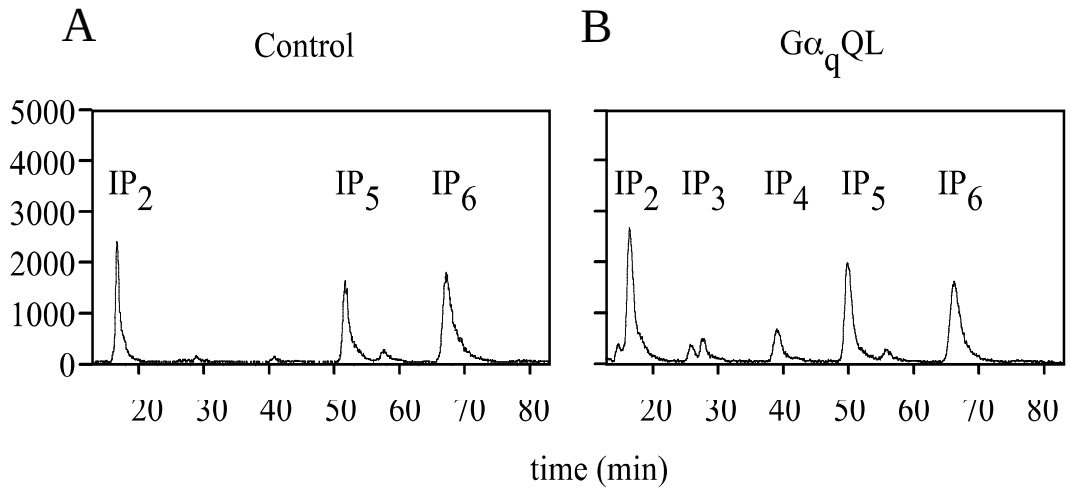


Figure 2.4. Stimulus dependent upregulation of the IP pathway. Adapted from Jim Otto et al PNAS 2007. Upon stimulus by $G\alpha_qQL$ Rat-1 cells metabolically labeled with $^3H$-Inositol experience a massive upregulation of the Inositol Phosphate pathway, but the highly phosphorylated $IP_5$ and $IP_6$ are largely unchanged. A) unstimulated and B) $G\alpha_qQL$ stimulated Inositol HPLC profiles. Copyright 2007 National Academy of Sciences

is not the case for $IP_6$. When Rat-1 cells are metabolically labeled with $^3H$ Inositol, and then the Inositol Phosphate pathway is upregulated by overexpression of $G\alpha_qQL$ there is a massive upregulation of the Inositol Phosphate pathway, except for $IP_6$ (as well as only mild upregulation of $IP_5$). The levels at resting state are fairly high, and in response to cellular stimulus, in this case expression of a constitutively active GPCR activator ($G\alpha_qQL$, a Gα subunit that is constitutively active and known to upregulate PLC dependent IP signaling), there is little change in $IP_5$ levels and almost no change in $IP_6$ levels (Otto *et al.*, 2007). This indicates that $IP_6$ may be acting in a way that is different from that of a classical second messenger. One possibility is that $IP_6$ is binding to proteins in a way that is less transient than a classical second messenger. It may be binding to proteins in a stable manner where it does not exchange with the bulk content of the cell. In this way $IP_6$ is not acting as a transient messenger, but rather as a structural cofactor that remains associated with the protein. There is indeed a growing number of proteins that fit within this model of $IP_6$ as a structurally relevant factor for proteins. Some of these

proteins that display a more stable structural interaction with $IP_6$ will be described below.

### 2.9.1 Adar2

The first example of a protein that binds to $IP_6$ in a structural and non-exchangeable way was Adar2, described above as being involved in RNA processing. This protein functions as an RNA editing protein that converts A to I and provides a differential splicing (Keegan, Gallo and O'Connell, 2001), and it had been observed that active Adar2 protein could not be purified from prokaryotic cells (Ring, O'Connell and Keegan, 2004), including the standard *E. coli* expression systems which lack Inositol Phosphates. In order to purify protein for RNA editing assays protein had to be produced in eukaryotic expression systems like the yeast *Pichia pastoris* (Ring, O'Connell and Keegan, 2004). When the crystal structure of Adar2 was solved in 2005, the protein was purified from *S. cerevisiae*, which do contain abundant $IP_6$, and interestingly $IP_6$ was observed in the crystal structure deeply buried within the protein as seen in Figure 2.5 (Macbeth, Schubert, VanDemark, Lingam, Hill, Bass, *et al.*, 2005). It was also demonstrated that in lysates

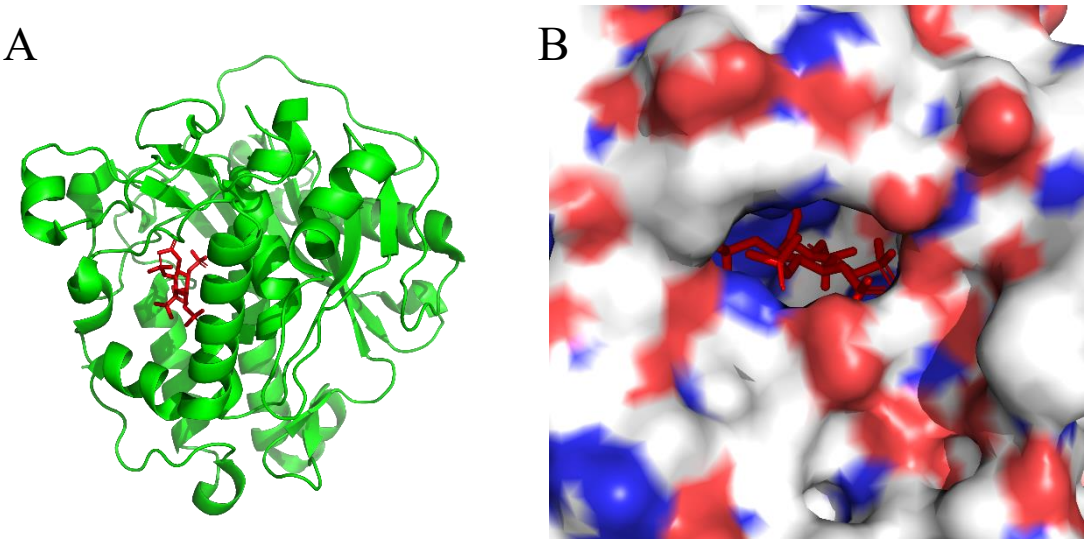


Figure 2.5. $IP_6$ bound to Adar2 (PDB 1ZY7). $IP_6$ is bound to Adar2 within a deep pocket of positively charged amino acids. A) Ribbon diagram of overall Adar2 structure and B) surface representation of $IP_6$ binding pocket.

from yeast lacking more highly phosphorylated IPs that the yeast homologue of Adar2, Adat1 was unable to perform RNA editing (Macbeth, Schubert, VanDemark, Lingam, Hill, Bass, *et al.*, 2005).

### 2.9.2 Plant Hormone Receptors

After Adar2 was identified as an $IP_6$ binding protein with the solving of its crystal structure, the structure of the plant hormone receptor Tir1 was solved, and $IP_6$ was found buried within the protein, this time within a Leucine Rich Repeat (LRR) repeat domain that formed a large super solenoid seen in Figure 2.6A-C (Tan *et al.*, 2007). The $IP_6$ was again found to be bound coordinated by many positively charged amino acids, and in this case IP binding was found to enhance the binding of the receptor to its ligand. Another plant hormone receptor closely related to Tir1 called Coi1 was identified as an IP binding protein when $IP_5$ was identified by Mass Spectrometry to co purify with Coi1. Coi1 has a similar positively charged binding pocket formed in the center of an LRR solenoid domain, and the binding of $IP_5$ has been demonstrated to enhance the affinity of Coi1 with its ligand as well (Sheard *et al.*, 2010).

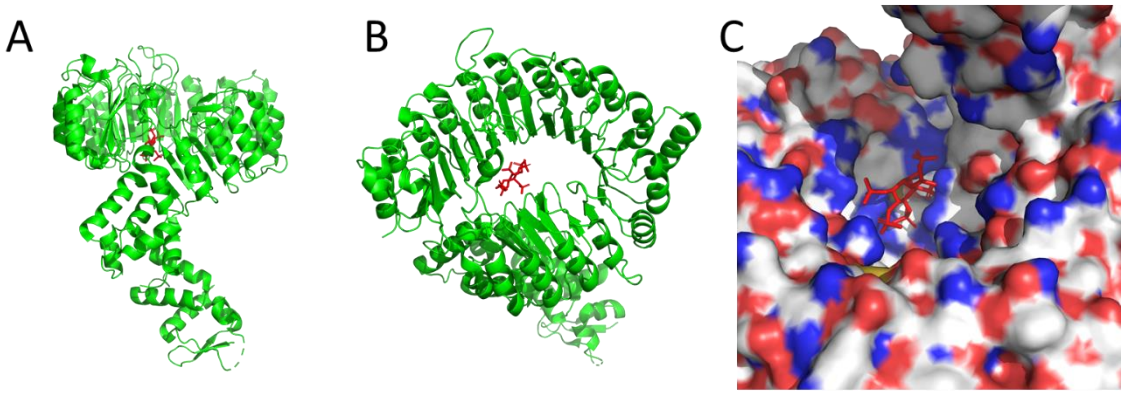


Figure 2.6. $IP_6$ bound to Tir1-Ask1 complex (PDB 2P1M and 3C6N). Ribbon diagram that $IP_6$ is bound to the plant hormone receptor Tir1 (A-B) and surface representation demonstrating IP6 binding mediated through interactions with a positively charged pocket within the center of a super solenoid formed by LRR repeat domains.

### 2.9.3 N-Terminal Acetylases

After these initial descriptions of proteins that seem to be interacting with $IP_6$ in more than a transient way, as well as the observation that $IP_6$ does not seem to behave in the same way as other second messengers, the York lab designed a screen to identify $IP_6$ binding proteins that have a much more stable interaction with $IP_6$. This screen was based around a kinase assay utilizing Vip1, a kinase that adds a pyrophosphate to $IP_6$ at the 1 position. When incubated with $IP_6$, and $^{32}P\gamma$ ATP, the Vip1 kinase will add the radioactive phosphate to $IP_6$ (Figure 2.7A), and then utilizing Thin Layer Chromatography the radiolabeled $IP_7$ can be detected (Figure 2.8). The key observation that allowed this to be utilized with the screen was that when a protein like

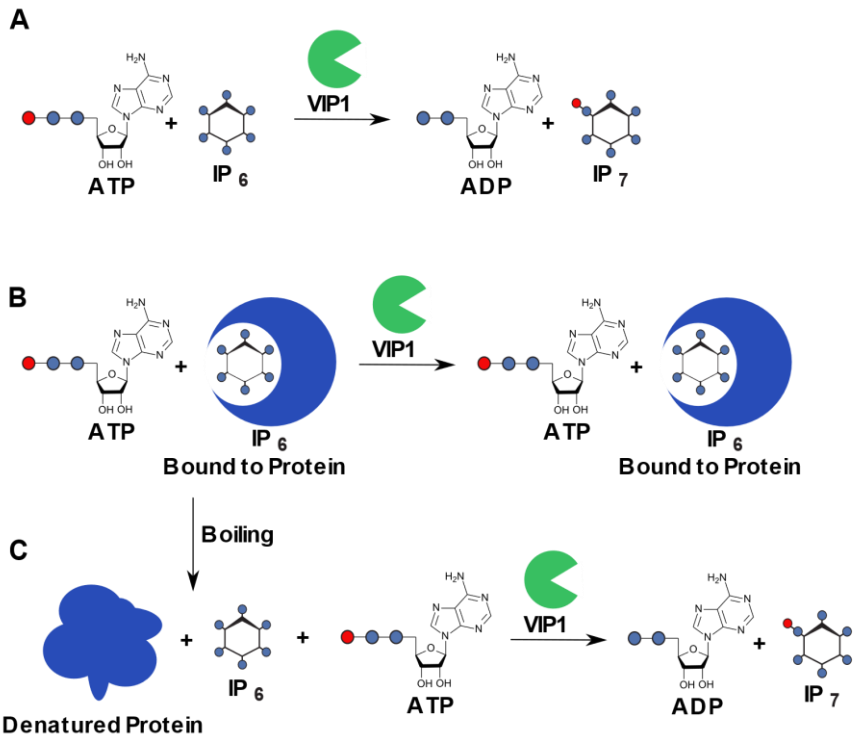


Figure 2.7. A boil dependent Vip1 kinase assay to identify $IP_6$ binding proteins. A) When Vip1 is incubated with $IP_6$ and $^{32}P\gamma$ ATP $^{32}P$-1-$IP_7$ is produced. B) When Vip1 is incubated with $^{32}P\gamma$ ATP and $IP_6$ that is bound structurally to a protein there is no formation of $IP_7$ since the $IP_6$ is not accessible to the kinase, however if the protein is first boiled C), the protein is denatured, and the $IP_6$ is released and $^{32}P$-1-$IP_7$ is formed. Adapted from Pham 2012.

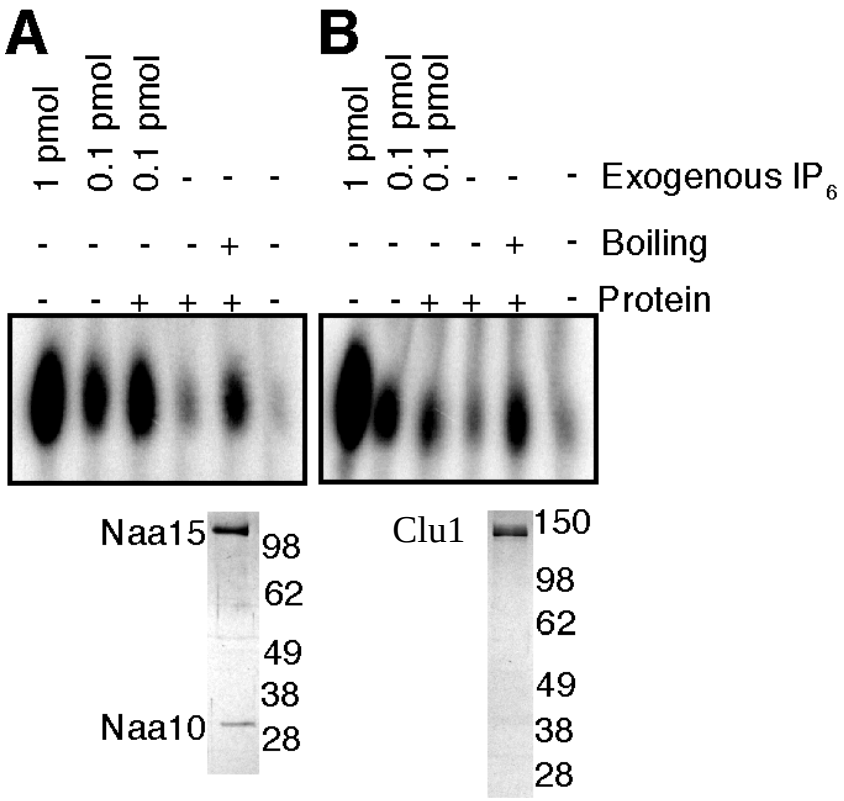


Figure 2.8. NatA and Clu1 are $IP_6$ binding proteins. The boil dependent kinase assay utilized as part of a screen to identify two new $IP_6$ binding proteins A) NatA and B) Clu1. Figure from Pham 2012. Copyright 2012 Trang Pham

Adar2, that tightly binds to $IP_6$ was incubated with Vip1 kinase, no $IP_7$ was produced since the $IP_6$ was bound within Adar2, and therefore not accessible to the kinase (Figure 2.7B). If the Adar2 was first denatured by boiling or otherwise, the $IP_6$ was released and thus available to the kinase and $IP_7$ was formed (Figure 2.7C). Utilizing this boil dependent kinase assay, members of the York lab identified two novel $IP_6$ binding proteins. One of them was NatA (Figure 2.8A), the N-terminal acetyltransferase that is responsible for modifying approximately 30% of all cellular proteins with an N-terminal acetyl group, and the other was Clu1 (Figure 2.8B), a protein required for mitochondrial distribution (Pham, 2012). The crystal structure of NatA was solved, and indeed $IP_6$ was found to be bound, seen in Figure 2.9 (Neubauer, 2012). $IP_6$ was found to stabilize the interaction between the two subunits of the NatA heterodimer Ard1 and Nat1 seen

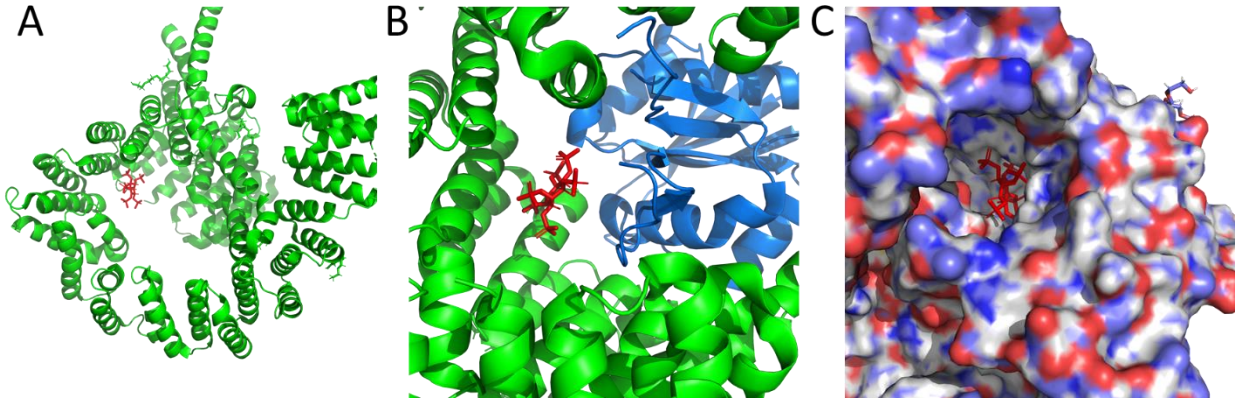


Figure 2.9. $IP_6$ bound to NatA (PDB 4HNW). NatA binds to $IP_6$ between its two subunits Nat1 and Ard1, at a positively charged pocket formed by the supersolenoid repeat formed by 14 TPR repeat domains of Nat1. A) Nat1 subunit with $IP_6$ bound at solenoid with Ard1 subunit not shown. B) $IP_6$ bound between the Nat1 and Ard1 subunits. C) NatA complex $IP_6$ binding pocket.

in Figure 2.9B. Interestingly, the $IP_6$ binding site was in the center of a large supersolenoid (Figure 2.9A and C) like the plant hormone receptors, however this time instead of a solenoid formed by LRR repeat domains, it was a solenoid of the Tetratricopeptide, or TPR repeat domains. Other N-Terminal acetylase complexes also were identified as $IP_6$ binding proteins including NatE and NatC (Pham, 2012), and the crystal structure of NatE (PDB-4XNH) also revealed that it too bound to $IP_6$, and that again, the interaction between the Ard1, Nat1 and Nat5 subunits were stabilized by binding of $IP_6$. This interaction was further characterized, and it was demonstrated that the human Nat complex was dependent on $IP_6$ for function, and that the yeast Nat complex required $IP_6$ for function under temperature stress conditions (Pham, 2012).

### 2.9.4 Pds5

Similar to the identification of the Tir1 and Adar2, another $IP_6$ binding protein, Pds5, a member of the chromatid cohesion complex named for the phenotype of precocious dissociation of sister chromatids, was found to bind to $IP_6$ when its crystal structure was solved as seen in Figure 2.10 (Ouyang *et al.*, 2016a). $IP_6$ was found to bind within a large supersolenoid domain, that was this time formed by HEAT repeats (named for the proteins where it was initially described Huntingtin, elongation factor 3 (EF3), protein phosphatase 2A (PP2A), and TOR1). When the residues responsible for $IP_6$ binding were mutated, the levels of protein are decreased, both in HeLa cells as well as in insect cell expression systems. The binding of $IP_6$ to Pds5 was shown to be required for its interactions with other members of the cohesion complex (Ouyang *et al.*, 2016a).

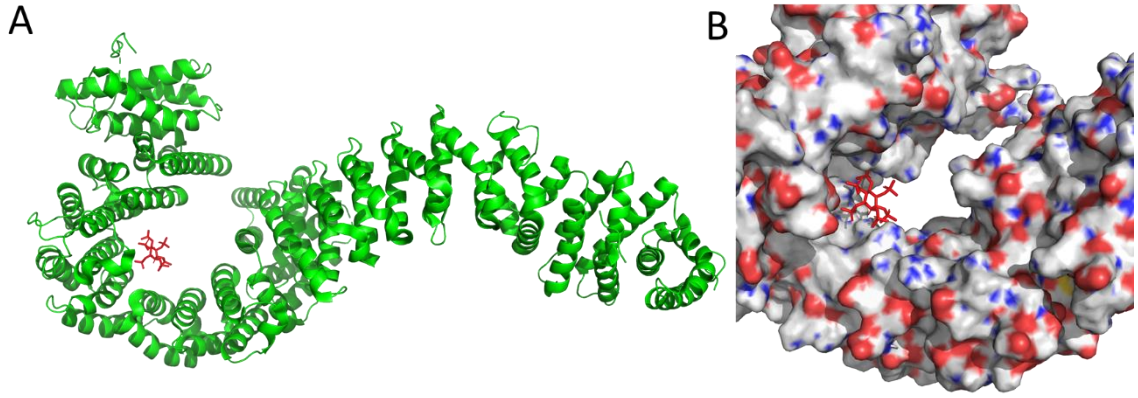


Figure 2.10. $IP_6$ bound to Pds5 (PDB 5HDT). Pds5 binds to $IP_6$ at a positively charged pocked formed at the center of a supersolenoid formed by HEAT repeat domains. A) Ribbon representation of $IP_6$ bound to Pds5 at a kink in the supersolenoid between HEAT repeats 15 and 16. B) Surface representation of $IP_6$ bound within Pds5 solenoid

### 2.9.5 Clu1

In the screen for $IP_6$ binding proteins described above, in addition to the NatA complex, another protein was also identified as a structural $IP_6$ binding protein, Clu1. Named for its

phenotype of clustered mitochondria observed across many species, little is known about the function of Clu1. Similar to many other $IP_6$ binding proteins, Clu1 was identified to contain repeat domains, a TPR domain like NatA, based on secondary structure prediction. It seems likely that Clu1 contains a supersolenoid and this coordinates its binding to $IP_6$. Little is known about the role of Clu1, and Chapter 4 of this work describes a characterization of Clu1, its binding to $IP_6$, and RNA, and the role that they play in what little is known of its function. Chapter 5 of this work is devoted to describing the structural characterization of Clu1, including Negative stain electron microscopy, Circular dichroism, in vitro refolding assays, Cryo electron microscopy, and X-Ray Crystallography.

## 2.10 Summary and Research Aims

Many protein structures have been determined with $IP_6$ as a ligand, and there is great diversity among both the functions of these proteins, and the nature of their interaction with $IP_6$. Some of these proteins are kinases and phosphatases that bind very specifically to IPs and add or remove a phosphate from a particular position, however there are also enzymes with looser specificity that recognize and modify multiple IPs. There are also proteins from bacteria and viruses that do not contain endogenous $IP_6$, but rather exploit eukaryotic cells' pool of $IP_6$ as a source of phosphate, a cofactor for activation of toxins, or as an assembly and stabilization factor. There are also several proteins that are involved in RNA processing that bind to $IP_6$ as well.

While early studies on many IPs were initially focused on their role as second messengers, this paradigm does not hold true for all of the IPs, $IP_6$ in particular does not behave as a second messenger, since it is present in cells at quite high concentrations, and upon stimulus

the levels do not change. This led the York lab to develop a screen for proteins that bind to $IP_6$ as a structural cofactor. In addition to the proteins identified with this screen several other proteins have been identified when $IP_6$ was found tightly bound inside of their crystal structures. Some of the members of the growing class of proteins that bind to $IP_6$ as a structural cofactor have been studied, but there are others that are less characterized, including one, Clu1, whose structure has yet to be solved. This dissertation describes the characterization of this $IP_6$ binding protein, both biochemically and structurally.

# Chapter 3 Reconstitution of Lipid Inositide Signaling in *E. coli*



## 3.1 Introduction

Myo-Inositol, a small cyclohexanol, is the precursor for a family of inositide molecules that control many diverse cellular and organismal functions (Berridge *et al.*, 1983; Raucher *et al.*, 2000; Otto *et al.*, 2007; Kantner, Schöll and Yanchuk, 2014). The first major class of inositol derivatives are the lipid phosphoinositide phosphates (PIPs), composed of phosphatidylinositol (PI) and its 7 phosphorylated derivatives. PIP signaling is initiated through production of PI, formed by Pis1 from cytidine diphosphate diacylglycerol (CDP-DAG) and myo-inositol. PI can then be further phosphorylated to form 3 PIPs: PI(3)P, PI(4)P and PI(5)P. These PIPs can then be further phosphorylated to form $PI(4,5)P_2$, the most abundant $PIP_2$, $PI(3,4)P_2$, and $PI(3,5)P_2$. $PI(4,5)P_2$ can then be phosphorylated by p110 (class-I PI3K in humans) to form $PI(3,4,5)P_3$, a key signal for cell growth and survival. Numerous studies of inositide lipids highlight their roles in uni- and multi-cellular eukaryotes (Fruman, Meyers and Cantley, 1998; Cantley, 2002; Irvine, 2016 Majerus and York, 2009; Balla, 2013).

In addition to lipid signaling, $PI(4,5)P_2$ is cleaved by phospholipase C (PLC) to form diacylglycerol and $IP_3$, which act as second messengers for protein kinase C activation and calcium release. Subsequently, $IP_3$ may be further phosphorylated to produce $IP_4$, $IP_5$, and $IP_6$, as well as pyro-phosphorylated to produce $IP_7$ and $IP_8$. These soluble inositol derivatives are conserved throughout eukaryotic organisms and are required for ion channel regulation, phosphate sensing, transcription, mRNA export, embryonic development, and act as structural cofactors (Hatch and York, 2010). A wide range of cellular and organismal roles have been defined for the collection of dozens of water soluble inositide messengers (Alcázar-Román and

Wente, 2008; Monserrate and York, 2010; Wilson, Livermore and Saiardi, 2013).

The complexity of inositide metabolism as well as redundancies in both kinase and phosphatase gene product families has clouded interpretation of enzyme specificity, regulation and inositide product function. Additionally, *in vitro* analyses, especially for lipid metabolizing enzymes, are challenging as they require recapitulation of the complex membrane, intermembrane and cofactor properties. As a means to address some of these issues, we initiated studies in bacteria because they lack endogenous or orthologous inositide signaling gene products. Our goal was to recapitulate simplified versions of both inositide lipid and soluble metabolic pathways. A previous study of heterologous expression of yeast phosphatidylinositol (PI) synthase in bacteria suggested the production of PI in prokaryotes demonstrated feasibility (Nikawa, Kodaki and Yamashita, 1988); however, expression of the extended inositide signaling pathway has yet to be reported. Here, we describe the development of a synthetic cell-based system through introduction of eukaryotic inositide lipase and kinase gene products into *E. coli*. Our system provides a greatly simplified platform that enables detailed and scalable cell-based characterization of inositide kinase, phosphatase, lipase, regulatory and effector proteins. Studies reported here may help provide new insights and clarity into inositide enzyme specificity and function.

## 3.2 Methods

### 3.2.1 Cloning

Standard PCR based cloning methods were used to introduce genes into the Duet vector system as previously described (Tolia and Joshua-Tor, 2006). Plasmids used in this study are listed in Table 3.1. For genes originating from *S. cerevisiae* or *B. subtilis*, genomic DNA was utilized for cloning. For genes from other species commercially available cDNA was utilized.

| Table 3.1: Plasmids Used | | |
|---|---|---|
| **Plasmid** | **Insert** | **Source** |
| **pET-duet scPis1** | *Saccharomyces cerevisiae* Phosphoinositide Synthase (sc Pis1) | (Clarke et al., 2019) |
| **pET-duet scPis1 btPik1** | *(sc Pis1), and Bos tarus* Phosphoinositide 4 kinase beta (bt Pik1) | (Clarke et al., 2019) |
| **pET-duet scPis1 scVPS34HELCAT** | (sc Pis1), and Saccharomyces cerevisiae Vacuolar protein sorting HELical and CATalytic subunit (sc Vps34 HELCAT) | (Clarke et al., 2019) |
| **pACYC-duet scMss4** | *Saccharomyces cerevisiae* Multicopy suppressor of Stt4 mutation (sc Mss4, a PI4P 5 kinase) | (Clarke et al., 2019) |
| **pACYC-duet scMss4 mmPLCd1** | (sc Mss4), and *Mus musculus* Phospholipase C delta 1 (mm Plcδ1) | (Clarke et al., 2019) |
| **pACYC-duet scMss4 hp110** | (sc Mss4), and *Homo sapiens* Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha (hs p110) | (Clarke et al., 2019) |
| **pCOLA-duet atIpk2** | *Arabidopsis thaliana* Inositol phosphate kinase 2 (at Ipk2) | (Clarke et al., 2019) |
| **pCOLA-duet atIpk2 scIpk1** | (at Ipk2), and *Saccharomyces cerevisiae* inositol phosphate kinase (sc Ipk1) | (Clarke et al., 2019) |
| **cup1-PLC1** | yeast Expression vector with scPlc1, copper inducible | (Stevenson-Paulik et al., 2006) |
| **pET24a PIP2 Operon** | Operon expression system for scPis1 btPik1 and scMss4 | (Clarke et al., 2019) |
| **pET21a IP Operon** | Operon expression system for mmPlcδ1 atIpk2 and scIpk1 | This study |
| **pET24a Full Operon** | Operon expression system for scPis1 btPik1 scMss4 mmPlcδ1 atIpk2 and scIpk1 | This study |
| **pBAD-Cam-Full Operon** | Operon expression system for scPis1 btPik1 scMss4 mmPlcδ1 atIpk2 and scIpk1 under an arabinose inducible promoter | This study |
| **pET28 scFab1** | *Saccharomyces cerevisiae* Forms Aploid and Binucleate cells (sc Fab1) | This study |
| **pET28 scPik1** | *Saccharomyces cerevisiae* Phosphatidyl Inositol Kinase (sc Pik1) | This study |
| **pET23 scPik1** | *Saccharomyces cerevisiae Phosphatidyl Inositol Kinase* (sc Pik1) | This study |
| **pET23 scStt4** | *Saccharomyces cerevisiae* STaurosporine and Temperature sensitive (sc Stt4) | This study |
| **pET23 scLsb6** | *Saccharomyces cerevisiae* Las Seventeen Binding protein (sc Lsb6) | This study |

The synthetic operons were designed as described below, and were synthesized, codon optimized, cloned and sequenced to verify the correct final sequence by Biomatik Corp. The resulting plasmids were combined and subcloned with standard restriction enzyme digestion and ligation. Plasmids in a variety of combinations were transformed into chemically-competent BL21 DE3 cells, BL21 Rosetta cells, or BL21 DE3 AI cells for plasmids with an Arabinose inducible promoter.

### 3.2.2 Growth of Bacteria

Bacterial strains expressing the Synthetic Inositide signaling cassette were grown as follows. First, cultures were grown in inositol-free M9 minimal salts with 1 mM Thiamine, 0.4% glycerol, 0.2% casamino acids, 2 mM $MgSO_4$ 0.1 mM $CaCl_2$ plus appropriate antibiotic selection at 37°C until they reached OD600 0.4.

### 3.2.3 Metabolic Labeling of Bacteria

To metabolically label bacteria they were grown in supplemented M9 media as described above and 100 µCi of $^3$H-inositol and 1mM Isopropyl b-D-1-thiogalactopyranoside (IPTG) were added (or 2% Arabinose for plasmids utilizing a pBAD promoter), and cells were grown at 30°C for 36 hours, with the addition of fresh antibiotics every 12 hours. Cells were then harvested by centrifugation and the cell pellets were washed with M9 minimal salts and stored at -80°C until use.

### 3.2.4 Extraction of PIPs

To extract Inositol Phosphates (IPs) and Phosphatidylinositol Phosphates (PIPs) cell pellets were resuspended in 100µl 0.5 M HCl, then 372 µl $CHCl_3$:MeOH 1:2, and 100 µl glass

beads were added and cells were lysed with bead-beating. Then, 125µl of each $CHCl_3$ and 2M KCl was added and the samples were subjected to additional bead-beating. After centrifugation at 14,000 RPM, the upper aqueous phase contained soluble IPs, and the lower organic phase contained the lipid PIPs. IPs in the upper phase were immediately analyzed with High Performance Liquid Chromatography (HPLC), while the lower organic phase containing PIPs were analyzed by thin—layer chromatography (TLC) or dried down with 10µl 100mg/ml bovine brain extract in $CHCl_3$:MeOH 50:1 carrier lipids, followed by deacylation prior to HPLC analysis separating gro-PIPs.

### 3.2.5 Deacylation of PIPs

Deacylation was conducted by adding 300µl 33% methylamine in ethanol:water:1-butanol 10:3:1 and incubated with rigorous agitation at 53°C for 1hr, followed by the addition of 150 µl of ice-cold 2-propanol and dried with a SpeedVac at 80°C. The dried pellet was resuspended in 200 µl water and extracted 3 times with 300 µl of 1-butanol:petroleum-ether:ethyl-formate (20:4:1), keeping the lower organic phase containing deacylated PIPs for HPLC analysis.

### 3.2.6 Preparation of Standards

W303 *S. cerevisiae* containing a plasmid expressing Plc1 with a Cup1 promoter (Stevenson-Paulik *et al.*, 2006) were grown in CSM and 50 µCi of $^3$H-inositol starting with 10 µl of overnight culture per ml of media. Cells were grown overnight at 30°C with 100µm $CuSO_4$, harvested by centrifugation, washed in PBS, then stored at -80°C until use. For hyperosmotic shock, wild-type yeast were grown as described above, before harvesting they were subjected to osmotic shock as previously described (Bonangelino *et al.*, 2002). For the PIP standards, mouse

embryonic fibroblasts (MEFs) were grown as previously described (Frederick *et al.*, 2005). Briefly, cells were grown in inositol-free DMEM supplemented with 10% FBS prior to labeling with 50 µCi of $^{3}$H-inositol for 3 days. Cells were washed with PBS, harvested in 100 µl 0.5M HCl, and immediately separated into aqueous (containing IPs) and organic (containing PIPs) fractions as described above.

### 3.2.7 HPLC of groPIPs

Both IPs and deacylated PIPs were loaded on to a Partisphere Sax 5 4.6 x 125mm column and eluted at 1 ml/min with ammonium phosphate (pH 3.5) gradients from 10 mM (buffer A) to 1.7 M (buffer B). For IPs separation was conducted as described in (Otto and York, 2010), and for deacylated PIPs the following gradient was used: 0% B for 5 minutes, then to 7% B over 6 minutes, followed by isocratic elution at 7% for 5 minutes, then a linear increase to 14% B over 15 minutes, to 60% B over 10 minutes, and finally to isocratic elution for 5 minutes. For IPs and the groPIPs from MEFs, 1ml fractions were collected and mixed with 6 ml each UlitimaFlowAP scintillation fluid. Radioactivity was measured with a Perkin Elmer liquid scintillation analyzer. For deacylated PIPs, radioactivity was measured with an in-line βRam detector, IN-US, with MonoFlow4 scintillation fluid from National Diagnostics at a 3:1 mixing ratio and using a 1 ml flow cell.

### 3.2.8 TLC

Borate thin layer Chromatography analysis of PIPs was conducted as described (Walsh, Caldwell and Majerus, 1991). Briefly, Silica gel 60 HPTLC plates were prepared with immersion in a CDTA solution for 10 min followed by baking at 100°C for 20 min. Dried PIP samples

were resuspended in $CHCl_3$:MeOH 2:1, spotted on the plates, then developed with methanol (75 mL), chloroform (60 mL), pyridine (45 mL), boric acid (12 g), water (7.5 mL), 88% formic acid (3 mL), BHT (2,4-Di-tert-butyl-4-methylphenol) (0.375 g), and ethoxyquin (75 µl). Radioactivity was detected using a Typhoon 9000 phosphor-imager. Representative spots for select PIPs were scraped from the silica gel plate with a razor blade and were deacylated as described above. The deacylated PIPs were analyzed with HPLC as described above with fraction collection and scintillation counting. Oxalate thin layer chromatography was conducted as described previously (Stolz *et al.*, 1998). Briefly, Silica gel 60 HPTLC plates were immersed in 54 mM potassium oxalate, 2mM EDTA in 47.5% ethanol, followed by drying for 1hr at 100°C. Dried PIP samples were resuspended in $CHCl_3$:MeOH 2:1, spotted on the plates, then developed with $CHCl_3$/Acetone/MeOH/Acetic acid/water (80:30:26:24:14) until the solvent reached the top of the plate, and radioactivity was detected as described above.

## 3.3 Results

### 3.3.1 Canonical Lipid Inositide Signaling

We aimed to use *E. coli*, which lack orthologous inositide gene products, to heterologously reconstitute abridged components of the yeast and metazoan inositide lipid and soluble metabolic pathways (Figure 3.1). Beginning with lipid pathways, we introduced yeast Pis1, which catalyzes the formation of phosphatidylinositol (PI) from CDP-DAG and myo-

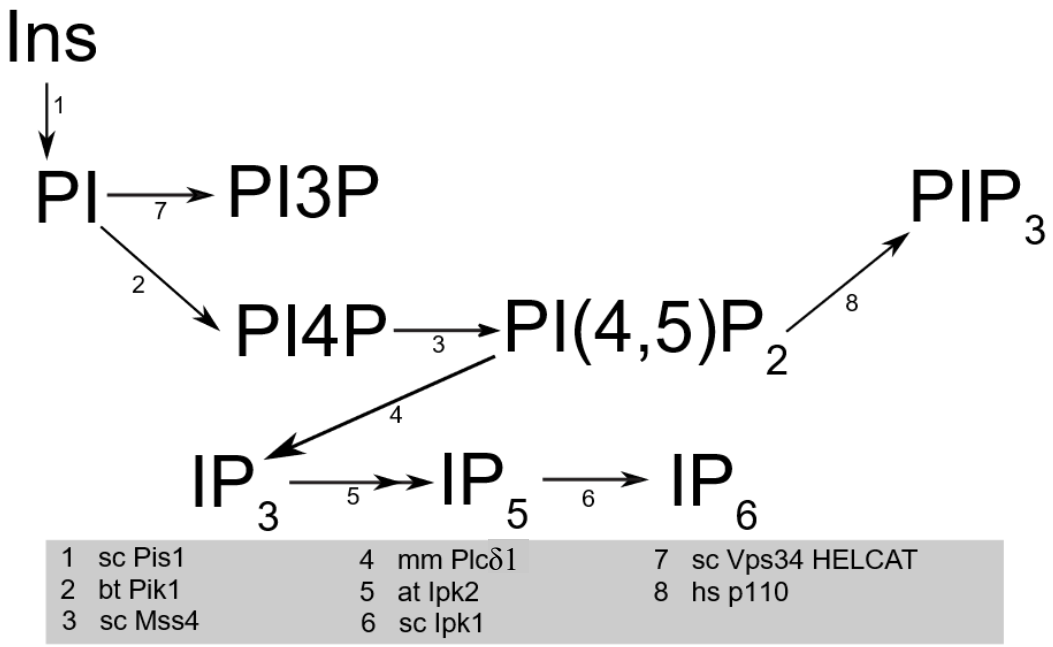


Figure 3.1. Synthetic approach to lipid derived inositide pathways in *E. coli*
The synthetic pathway used in this study to produce inositides in *E. coli*. Abbreviations: Myo-Inositol (Ins); Phosphatidyl Inositol (PI); Phosphatidyl Inositol 3-Phosphate (PI3P); Phosphatidyl Inositol 4-Phosphate (PI4P); Phosphatidyl Inositol 4, 5 Bisphosphate ($PI(4,5)P_2$); Phosphatidyl 3, 4, 5 Triphosphate ($PIP_3$), Inositol 1, 4, 5 Triphosphate ($IP_3$); Inositol 1, 3, 4, 5, 6 Pentakisphosphate ($IP_5$); Inositol Hexakisphosphate ($IP_6$); *Saccharomyces cerevisiae* Phosphoinositide Synthase (sc Pis1); *Bos tarus* Phosphoinositide 4 kinase beta (bt Pik1); *Saccharomyces cerevisiae* Multicopy suppressor of Stt4 mutation (sc Mss4, a PI4P 5 kinase); *Mus musculus* Phospholipase C delta 1 (mm Plcδ1); *Arabidopsis thaliana* Inositol phosphate kinase 2 (at Ipk2); *Saccharomyces cerevisiae* inositol phosphate kinase (sc Ipk1); *Saccharomyces cerevisiae* Vacuolar protein sorting HELical and CATalytic subunit (sc Vps34 HELCAT); *Homo sapiens* Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit α (hs p110).

inositol. In bacterial cells labeled with $^3$H-inositol, we observed the formation of PI as determined by comigration with standards on oxalate thin-layer chromatography (TLC) (Figure 3.2A lane 1). Next, as a means to generate phosphatidylinositol 4-phosphate, we expressed Pik1, a PI 4-kinase, and observed an additional lipid species consistent with PIP (Figure 3.2A lane 2). Similarly, to catalyze production of PI(3)P, we co-expressed Pis1 and Vps34, a PI-specific 3-

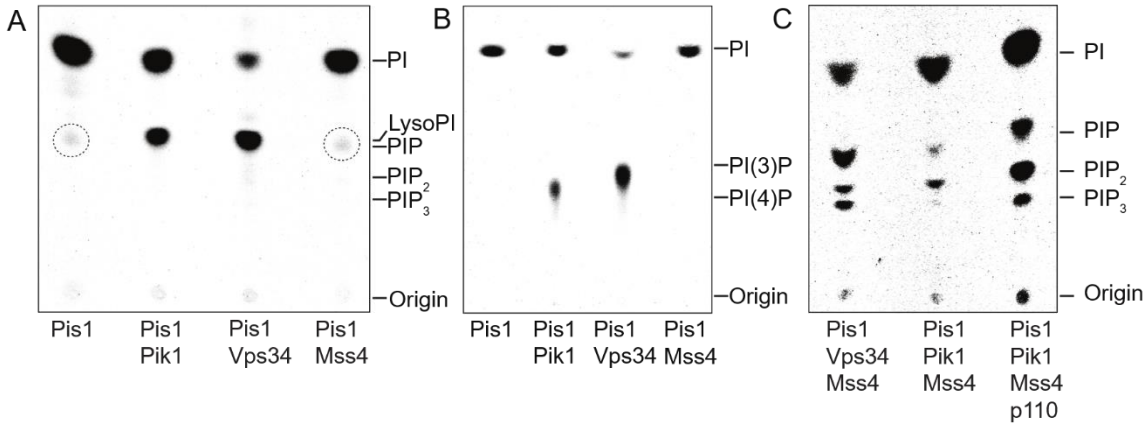


Figure 3.2. TLC analysis of reconstituted lipid inositide synthesis
Thin layer chromatography (TLC) analysis of 3H-Inositol labeled lipids. (A) Oxalate TLC showing Pis1 bacteria produce PI and Lyso-PI (dashed circle), Pis1-Pik1 bacteria produce PI and PIP, Pis1-Vps34 bacteria produce PI and PIP, and Pis1-Mss4 bacteria produce PI and Lyso-PI (dashed circle). E. coli expressing PI synthase Pis1 can produce PI, but also have some Lyso-PI, presumably from the activity of bacterial phospholipase A. (B) Borate TLC resolving PI(4)P and PI(3)P production by Pis1-Pik1 and Pis1-Vps34 bacteria, respectively, as well as resolution of lyso-PI from PIPs; (C) Oxalate TLC demonstrates that Pis1-Vps34-Mss4 bacteria produce PI, PIP, $PIP_2$ and $PIP_3$; Pis1-Pik1-Mss4 bacteria produce PI, PIP, $PIP_2$, and trace amounts of $PIP_3$; Pis1-Pik1-Mss4-p110 (PIK3CA)-expressing bacteria produce PI, PIP, $PIP_2$ and $PIP_3$.

kinase, revealing a robust accumulation of PIP and relative depletion of PI as compared to Pis1 alone (Figure 3.2A lane 3 vs. lane 1). Also, phosphorylation of PI by Vps34 was in the absence of Vps15, which is required for activation of Vps34 activity (Stack *et al.*, 1993). The robust activity of Vps34 compared to Pik1, and its ability to phosphorylate PI in the absence of Vps15, are consistent with published reports of the increased catalytic function of the truncated form of Vps34, which contains only the HELical and CATalytic domains, but not the regulatory domains, known as HELCAT (Miller *et al.*, 2010), that we introduced herein.

As a presumptive negative control for phosphorylation of PI, we expressed Pis1 along with Mss4, a PI(4)P 5-Kinase that is not reported to utilize PI as a substrate, and did not observe significant changes in lipid profiles as compared to Pis1 alone (Figure 3.2A, lane 4). The lipid species observed were characterized further by deacylation to their corresponding glycerol phosphoinositols and separation by high-performance chromatography (HPLC). This approach confirmed the identity of PI and revealed that the minor lipid, observed in lanes 1 and 4 that co-migrates with PIP standard, is lyso-PI, as the deacylated product was determined to be glycerol 1-phosphoinositol (Figure 3.2B, Figure 3.3C and 3.3F). We hypothesize, but do not prove, that the production of lyso-PI from the PI product of Pis1 may be a result of endogenous bacterial phospholipase A activities.

To further characterize and confirm the identities of the lipids produced in our bacterial system, we utilized two independent approaches: 1) a borate TLC system capable of differentiating phosphorylation at the D-3 and D-4 ring positions (Walsh, Caldwell and Majerus, 1991); and 2) deacylation of the glycerol-based bacterial lipids using methylamine, which then allows for high-resolution HPLC separation of the stereomers of the water soluble glycero-phosphoinositol (groPIP) products. Examination of radiolabeled lipid products using borate TLC indicates clear resolution of the Pis1-Pik1 produced PI(4)P (Figure 3.2B, lane 2) and Pis1-Vps34 produced PI(3)P (Figure 3.2B, lane 3). In addition, separation using the borate system improved resolution of lipid species as the lyso-PI observed in Pis1 and Pis1-Mss4 expressing cells no longer co-migrated with PIP (Figure 3.2B, lanes 1 and 4).

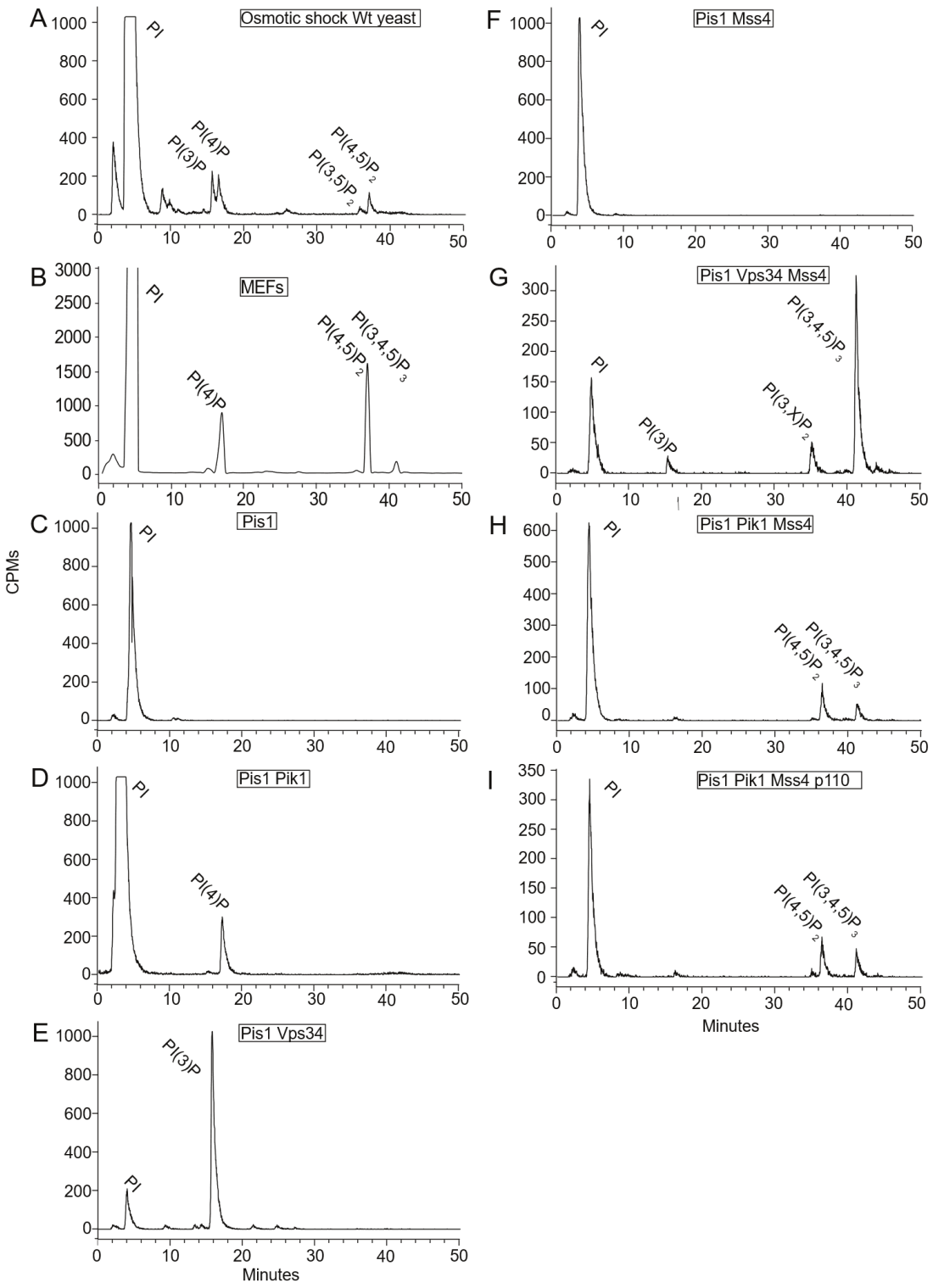
A
Osmotic shock Wt yeast
PI
PI(4)P
PI(3)P
PI(4,5)P2
PI(3,5)P2
B
MEFs
PI
PI(4)P
PI(4,5)P2
PI(3,4,5)P3
C
Pis1
PI
D
Pis1 Pik1
PI
PI(4)P
E
Pis1 Vps34
PI(3)P
PI
F
Pis1 Mss4
PI
G
Pis1 Vps34 Mss4
PI
PI(3)P
PI(3,X)P2
PI(3,4,5)P3
H
Pis1 Pik1 Mss4
PI
PI(4,5)P2
PI(3,4,5)P3
I
Pis1 Pik1 Mss4 p110
PI
PI(4,5)P2
PI(3,4,5)P3
CPMs
Minutes

Figure 3.3. HPLC analysis of reconstituted lipid inositide synthesis
Representative HPLC traces from analysis of 3H-inositol labeled glycerol-inositol phosphates from deacylated PIPs. (A) Osmotically shocked wild-type yeast showing separation of gro-PI(3)P and gro-PI(4)P, and of gro-PI(3,5)$P_2$ and gro-PI(4,5)$P_2$. (B) Mouse embryonic fibroblasts (MEFs) showing relative elution positions of gro-PI(4)P, gro-PI(4,5)$P_2$, and gro-PI(3,4,5)$P_3$. (C) Pis1-expressing bacteria produce of gro-PI (D) Pis1-Pik1 bacteria produce of gro-PI and gro-PI(4)P. (E) Pis1-Vps34 bacteria have depletion of gro-PI and robust production of gro-PI(3)P. (F) Pis1-Mss4 bacteria showing inability of Mss4 to phosphorylate PI. (G) Pis1-Vps34-Mss4 bacteria convert gro-PI to gro-$PIP_3$ with gro-PI(3)P and gro-PI(3,X)$P_2$ intermediates. Despite its traditional role as a PI(4)P 5-kinase, Mss4 has activity against PI(3)P, and the resulting PI(3,X)$P_2$ is converted very efficiently to $PIP_3$, either by Mss4, or possibly by Vps34. (H) Pis1-Pik1-Mss4 bacteria produce PI(4,5)$P_2$ thus demonstrating the canonical role of Mss4, but surprisingly also produce PI(3,4,5)$P_3$ without expression of a separate PI(4,5)$P_2$ 3-Kinase. (I) Pis1-Pik1-Mss4-p110 (PIK3CA) bacteria produce PI(4,5)$P_2$ and PI(3,4,5)$P_3$.

Equally important in the identification of lipid products is the use of deacylation and separation of the resulting glycerol-inositol phosphates utilizing HPLC. Glycerol-inositol phosphate standards were prepared from osmotically shocked $^3$H-inositol radiolabeled yeast (Figure 3.3A) or radiolabeled mouse embryonic fibroblasts (MEFs) (Figure 3.3B). Total lipid deacylation of bacteria confirmed that the Pis1 bacteria produced a single species coeluting with glycerol-inositol 1-phosphate (Figure 3.3C). Addition of either Pik1 or Vps34 along with Pis1 resulted in the appearance of glycerol-inositol 1-phosphate and an additional species co-eluting with either glycerol-inositol 1,4-phosphate or glycerol-inositol 1,3-phosphate (Figures 3.3D and E). This method also confirmed that Mss4 is unable to phosphorylate PI, as a single glycerol-inositol 1-phosphate species was observed in Pis1-Mss4 bacteria (Figure 3.3F).

While Mss4 is classically considered to be a PI(4)P specific 5-Kinase, moonlighting activities such as PI(3)P 4-kinase activity (Desrivières *et al.*, 1998) and PI(3,4)$P_2$ 5-kinase activity (Zhang *et al.*, 1997) have also been described. As a presumed negative control, we expressed Mss4 along with Vps34 and Pis1 and separated lipids by TLC (Figure 3.2C) and by deacylation/HPLC (Figure 3.3E). Unexpectedly, we observed the appearance of two new species and the concomitant reduction of PI(3)P (Figure 3.2C, lane 1). These data suggest that PI(3)P is converted to a PI(3,X)$P_2$ species, likely PI(3,4)$P_2$ based on previous reports (Zhang *et al.*, 1997; Desrivières *et al.*, 1998; Mitra *et al.*, 2004). Secondly, we observed the appearance of a substantial amount of a species co-migrating with $PIP_3$ (Figure 3.2C lane 1). These data were supported by analysis of the bulk deacylated Bligh-Dyer extracted radiolabeled lipids in which we observed the appearance of four species co-eluting with glycerol-inositol 1-phosphate, glycerol-inositol 1,3-bisphosphate, glycerol-inositol 1,3,x-trisphosphate and glycerol-inositol 1,3,4,5-tetrakisphosphate, respectively (Figure 3.3G).

Next, when Mss4 was co-expressed with Pis1-Pik1, we observed the expected depletion of PI(4)P and appearance of a species comigrating with PI(4,5)$P_2$, consistent with its canonical role as a PI(4)P 5-Kinase (Figure 3.2C lane 2). In addition, we report that Mss4 expressed in a bacterial system enables the unexpected formation of $PIP_3$, presumably a relatively low level of $PIP_2$ 3-kinase moonlighting activity, which has not previously been observed for Mss4 (Figure 3.2C lane 2). Additional analysis of these bacterial lipids by deacylation confirmed these activities by the observation of four (4) species co-eluting with glycerol-inositol 1-phosphate, glycerol-inositol 1,4-bisphosphate, glycerol-inositol 1,4,5-trisphosphate and glycerol-inositol 1,3,4,5-tetrakisphosphate, respectively (Figure 3.3H). As a positive control for $PIP_3$ production, we also co-expressed a class III $PIP_2$ 3-kinase catalytic subunit, p110, and observed additional

conversion of $PIP_2$ to $PIP_3$ by TLC (Figure 3.2C lane 3) and deacylated products (Figure 3.3I). Overall, our data demonstrate that Mss4 is capable of acting as a promiscuous PIP multikinase in both a PI(3)P- and PI(4)P-dependent manner to produce $PIP_2$ and $PIP_3$ species.

### 3.3.2 Utilizing the Bacterial System to Characterize Fab1

Next, we sought to utilize the bacterial system, and its lack of endogenous Inositide signaling to characterize the activity of the PI3P 5 kinase Fab1. In order to synthesize PI5P from PI via the canonically described pathway it must first be phosphorylated by Vps34 to produce PI3P, which then can be phosphorylated to Fab1 at the 5 position to produce $PI(3,5)P_2$. This can then be dephosphorylated by Sac or Inp5 to finally produce PI5P. While this is the primary pathway observed, there are also groups that claim Fab1 can act directly on PI to form PI5P without going through a $PI(3,5)P_2$ intermediate. The existence of this pathway has been difficult to demonstrate due to the limitations of *in vitro* assays on lipid kinases, as well as the convoluted compartment specific nature of inositol signaling, and the relatively low abundance of the lipids involved. These factors combined make this question an attractive target for study utilizing the bacterial system for inositide expression. In order to test the ability of Fab1 to phosphorylate PI, we co expressed Fab1 along with Pis1. These bacteria were indeed able to produce a small amount of PIP (Figure 3.4 lane 2), consistent with Fab1 having some degree of PI 5 kinase activity. As a positive control for the activity of Fab1, we also expressed Fab1 along with Pis1 and Vps34, where it would be expected to produce $PI(3,5)P_2$. Surprisingly this combination of genes produced $PIP_3$ (Figure 3.4 lane 2), which was confirmed by comigration with $PIP_3$

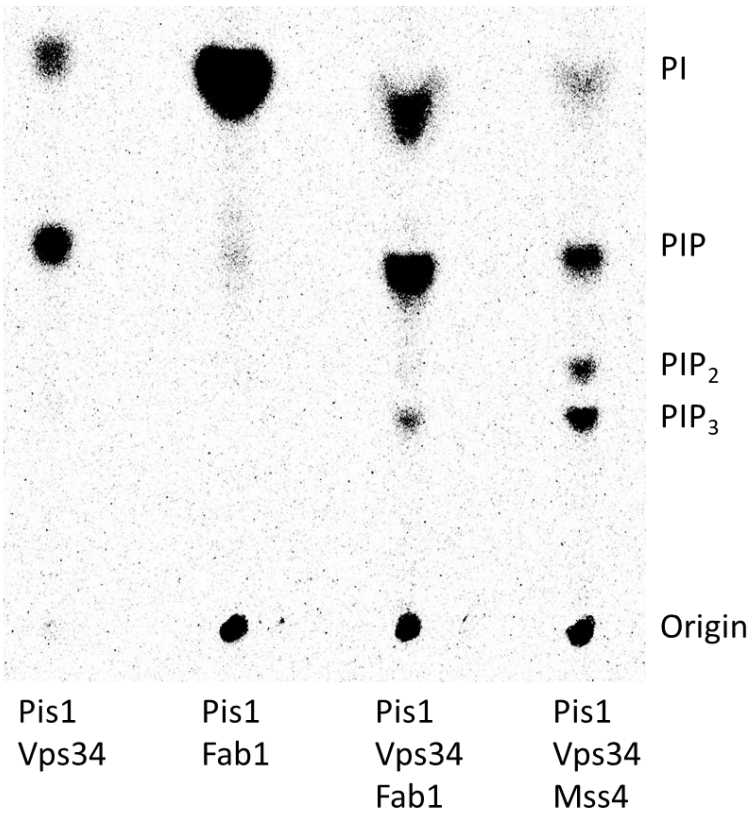


Figure 3.4 Fab1 can act as a PI kinase and produce $PIP_3$ with Vps34. Lane 1, as a positive control Pis1 and Vps34 together produce PI and PIP in bacteria. Bacteria expressing Pis1 and Fab1 produce a lot of PI, and trace amounts of PIP. When Vps34 is added $PIP_3$ is produced with trace amounts of $PIP_2$. As a positive control Pis1, Vps34 and Mss4 expressed together produce PI, PIP, $PIP_2$ and $PIP_3$.

produced by Pis1, Vps34 and Mss4 (Figure 3.4 lane 3). This observation could be explained either by Fab1 or Vps34 HELCAT also having some PI(3,5)$P_2$ 4 kinase activity. Since $PIP_3$ formation is also observed in bacteria expressing Pis1, Vps34 and Mss4, it could be argued that the Vps34 HELCAT construct may be exhibiting promiscuous activity in both cases leading to the formation of $PIP_3$ via two separate pathways. It could also be that in each of the different combinations the 5 kinases Mss4 and Fab1 each have some promiscuity, and this activity is responsible for the $PIP_3$ formation. More analysis of these bacterially produced inositides, including HPLC of deacylated PIPs utilizing tandem column columns as described by Sarkes and

Rameh, 2010 capable of resolving different PIP and $PIP_2$ species, is required to definitively confirm these activities. Further study of these enzymes in the bacterial context may lead to a better understanding of their activities.

### 3.3.3 Bacterial Reconstitution of the Soluble IP Pathway

As a means to study the lipid-derived water-soluble inositol phosphate pathways, we used our heterologous bacterial system to examine the activities of phosphoinositide selective phospholipase C (PLC) and inositol phosphate kinase (IPK) enzymes. To this end, we utilized Plcδ1 from mouse, a δ-class enzyme capable of cleaving $PI(4,5)P_2$ to produce DAG and inositol 1,4,5-trisphosphate, $I(1,4,5)P_3$. HPLC separation of $^3$H-inositol labeled water-soluble extracts from bacterial cells were compared to IPs produced in *S. cerevisiae* overexpressing Plc (Figure 3.5A). In order to characterize the ability of Plc to cleave lipids in the bacterial system, we expressed Plc along with bacterial strains also expressing the genes for each stage of the lipid inositide pathway starting with Pik1, and we observed no peaks corresponding with the IP standards (Figure 3.5B). When Plc was co expressed with Pis1-Pik1 we observed a peak that co-elutes with $I(1,4)P_2$, indicating that Plc is cleaving PI(4)P (Figure 3.5C). While bacteria do not endogenously produce IPs, they do contain a gene with homology to a mammalian $I(1,4)P_2$ 1-phosphatase called CysQ (Neuwald *et al.*, 1992). (Spiegelberg *et al.*, 1999). Here we demonstrate that *in vivo* CysQ 1-phosphatase activity is minimal, as $I(1,4)P_2$ is present (Figure 3.5C), however further characterization of this enzyme *in vivo* may prove of interest.

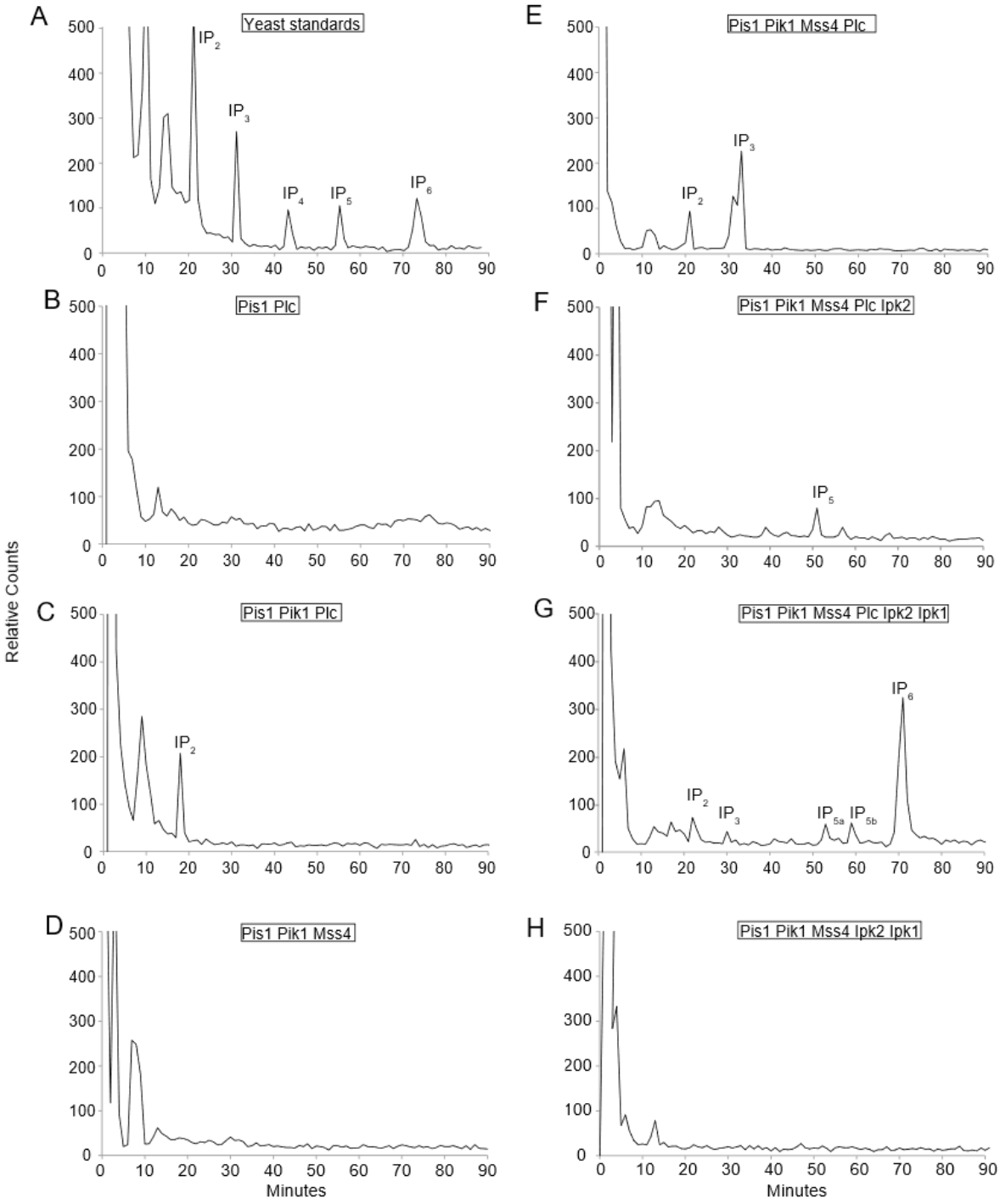
A
Yeast standards
IP2
IP3
IP4
IP5
IP6
B
Pis1 Plc
C
Pis1 Pik1 Plc
IP2
D
Pis1 Pik1 Mss4
E
Pis1 Pik1 Mss4 Plc
IP2
IP3
F
Pis1 Pik1 Mss4 Plc Ipk2
IP5
G
Pis1 Pik1 Mss4 Plc Ipk2 Ipk1
IP2
IP3
IP5a
IP5b
IP6
H
Pis1 Pik1 Mss4 Ipk2 Ipk1
Relative Counts
Minutes
Minutes

Figure 3.5. Reconstitution of soluble inositol phosphate synthesis. Representative traces from HPLC analysis of 3H inositol labeled soluble IPs. (A) S. cerevisiae overexpressing Plc1 produce I(1,4)$P_2$, I(1,4,5)$P_3$, I(1,4,5,6)$P_4$, I(1,3,4,5,6)$P_5$ and $IP_6$. (B) Pis1-Plc bacteria do not produce IPs due to a lack of substrate lipids for Plc. (C) Pis1-Pik1-Plc bacteria produce I(4,5)$P_2$, consistent with production of PI(4)P and relative lack of endogenous I(1,4)$P_2$ 1-phosphatase activity by CysQ. (D) Pis1-Pik1-Mss4 bacteria do not produce IPs. In the absence of Plc, inositide lipids and are not converted to soluble IPs. (E) Pis1-Pik1-Mss4-Plc bacteria produce I(1,4)$P_2$ and I(1,4,5)$P_3$. When Plc is added to bacteria producing PI(4)P and PI(4,5)$P_2$, these lipids are cleaved to produce their corresponding soluble IPs. (F) Pis1-Pik1-Mss4-Plc-Ipk2 bacteria produce I(1,3,4,5,6)$P_5$. When the IP multikinase Ipk2 is introduced into bacteria, it converts $IP_3$ to $IP_5$ consistent with its documented role across eukaryotes. (G) Pis1-Pik1-Mss4-Plc-Ipk2-Ipk1 bacteria produce I(1,3,4,5,6)$P_5$($IP_5$A), $IP_6$, as well as $IP_5$B. Ipk1 phosphorylates $IP_5$A to make $IP_6$, however the presence of $IP_5$B may be due to the activity of bacterial phytases. (H) Pis1-Pik1-Mss4-Ipk2-Ipk1 bacteria do not produce IPs, demonstrating that the soluble IP production is dependent on the action of Plc on select lipid inositide species.

In the absence of Plc, bacteria expressing Pis1-Pik1-Mss4 produce PI(4,5)$P_2$ (Figure 3.3H), but do not produce any soluble IPs (Figure 3.5D). When Plc is expressed along with Pis1-Pik1-Mss4 species that comigrate with I(1,4,5)$P_3$ and I(1,4)$P_2$ are produced, indicating that the PI(4)P and PI(4,5)$P_2$ are being cleaved by Plc (Figure 3.5E). The $IP_3$ peak present in these bacteria is wider than the $IP_3$ peak present in *S. cerevisiae* produced standards, and this may be due to the activities of bacterial phospholipases producing glycerol-inositol phosphates. Next, to test the ability of the IP multikinase Ipk2 to phosphorylate $IP_3$, and sequentially $IP_4$ to form $IP_5$, we expressed Ipk2 from *Arabidopsis thaliana*, along with Pis1-Pik1-Mss4-Plc, and a species coeluting with $IP_5$ was observed (Figure 3.5F). No peak comigrating with $IP_4$ is observed in this strain, indicating that the two-step Ipk2 catalyzed conversion of $IP_3$ to $IP_5$ precedes in this bacterial system without accumulation of an $IP_4$ intermediate. Of note, expression of the lipid inositide pathway, Plc, and the soluble IP kinases required the utilization of multiple plasmids, multiple antibiotics, and combined with the decreased growth rate in inositol-free M9 minimal media, the incorporation of $^3$H-inositol label was dramatically lower in this strain, potentially

limiting our ability to detect minor species of IPs, like an $IP_4$ intermediate. Finally, to form $IP_6$, we expressed yeast Ipk1, which phosphorylates I(1,3,4,5,6)$P_5$ at the 2-position. When Ipk1 is expressed along with Pis1-Pik1-Mss4-Plc1-Ipk2, a peak coeluting with $IP_6$ is observed, along with peaks corresponding with two different $IP_5$ species (Figure 3.5G). We hypothesize, but do not prove, that the first $IP_5$ peak to elute ($IP_{5A}$) is I(1,3,4,5,6)$P_5$, the product of Ipk2, and the substrate of Ipk1, and the second ($IP_{5B}$) is the result of a bacterial phytase present in *E. coli* (Greiner, Konietzny and Jany, 1993), which displays promiscuous phosphatase activity, acting to remove a phosphate from $IP_6$ produced by Ipk1. To demonstrate that this inositide pathway is lipid-dependent, we expressed Pis1-Pik1-Mss4-Ipk2-Ipk1, but without Plc, and expectedly do not observe production of any IP species (Figure 3.5H).

### 3.3.4 An Operon-Based Approach to Inositide Signaling

One of the problems encountered with the expression of the phosphoinositide pathway in bacteria as described thus far is the use of multiple plasmids and the difficulty of maintaining as many as 3 or 4 plasmids within the same bacteria. Growth in the presence of this many antibiotics results in very poor growth of the bacteria, and has been a serious limit to the utility of the system, as well as limiting the applications, since in order to test the impact of other kinases, phosphatases, or inositide binding proteins even more plasmids will need to be utilized. In order to address this limitation a synthetic operon was designed to co express larger combinations of genes all under the same promoter, in a manner similar to how endogenous cassettes of genes are expressed in bacteria. These large sequences were codon optimized for expression in *E. coli*, and were synthesized, and verified by Biomatik. The operon was designed to be synthesized into two vectors. The first contained the PI synthase scPis1, the PI 4 kinase

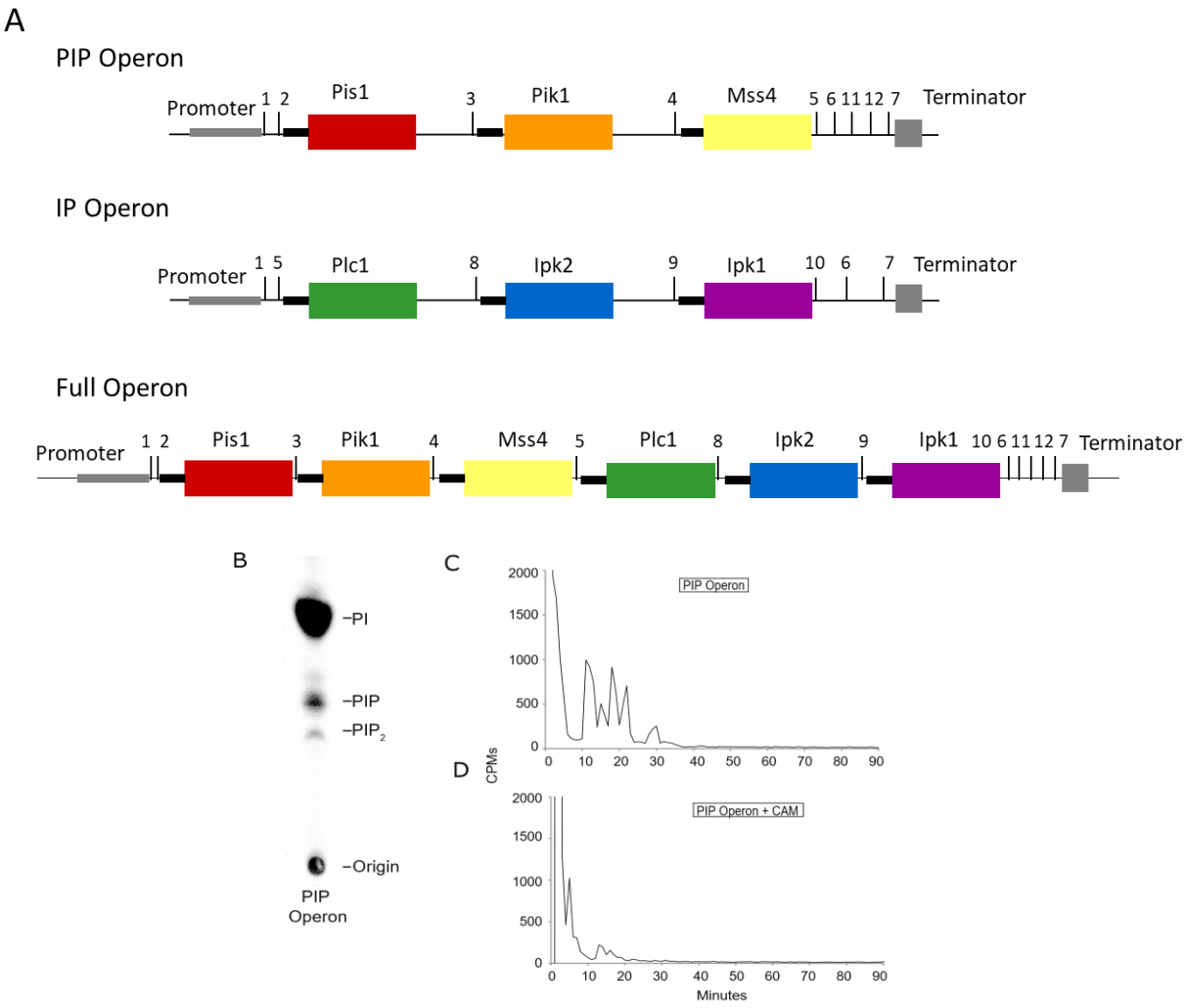


Figure 3.6. A synthetic operon to express the inositide pathway in *E. coli* (A). Codon optimized sequence for scPis1, btPik1, and scMss4 were each flanked by restriction enzyme sites and preceded by the spacer sequence 5'-AATAATTTTGTTTAAC-3' and the RBS 5'-AGGAGGTATATATA-3' and the entire construct was inserted between the XhoI and XbaI sites in pET24a. All three genes in the operon are under the control of the T7/Lac promoter and the T7 terminator. Oxalate TLC separation (B) of extracts from BL21 Rosetta cells expressing the PIP operon demonstrates the production of PI, PI(4)P, and PI(4,5)$P_2$. HPLC separation of the aqueous phase from extracts prepared from bacteria expressing the PIP operon without (C) or with Chloramphenicol (D). In bacteria expressing the PIP operon without Chloramphenicol, there are several peaks present in the aqueous phase (C). These may be due to the activity of bacterial phospholipases on PI(4,5)$P_2$ that are in inhibited by Chloramphenicol since these peaks are absence with the PIP operon is expressed with Chloramphenicol.

| Table 3.2 Restriction Enzyme Sites in the Inositide Operon | | |
|---|---|---|
| **Site** | **Enzyme** | **Feature** |
| **1** | XbaI | Entire Cassette |
| **2** | EcoRI | Pis1 |
| **3** | HindIII | Pis1-Pik1 |
| **4** | KpnI | Pik1-Mss4 |
| **5** | BamHI | Mss4-MCS-Plc1 |
| **6** | NheI | MCS |
| **7** | XhoI | MCS-Entire Cassette |
| **8** | NdeI | Plc1-Ipk2 |
| **9** | XmaI | Ipk2-Ipk1 |
| **10** | NcoI | Ipk1-MCS |
| **11** | PstI | MCS |
| **12** | SacI | MCS |
| **For cloning of Cassette 2 (the IP operon) into Cassette 1 to produce the full operon sites 5 and 6 are utilized.** | | |

btPik1, and the PI4P 5 kinase Mss4, as well as a multicloning site to allow for expansion of the operon to include additional genes as desired. The second operon would allow for the expression of the three genes required to convert $PI(4,5)P_2$ to $IP_6$; mmPLCδ1 to cleave $PIP_2$ and produce the soluble $I(1,4,5)P_3$, next atIPMK to convert the $IP_3$ first to $IP_4$ and then $IP_5$, and finally the scIpk1 to convert the $IP_5$ to $IP_6$. The second operon was also designed to be compatible with the first, both for co expression, but also to be cloned into the multicloning site of the first operon, to all the simultaneous expression of all 6 genes required for the lipid dependent $IP_6$ synthesis pathway. A schematic of the operon is shown in Figure 3.6A, and an overview of the restriction sites is given in Table 3.2.

First, we sought to characterize the lipid operon designed to produce $PI(4,5)P_2$, by expressing Pis1, Pik1, and Mss4. We confirmed utilizing TLC that bacteria expressing these three genes in a synthetic operon were able to produce PI, PI4P and $PI(4,5)P_2$ (Figure 3.6B). We also noted that relative to expression with the Duet system, bacteria expressing the synthetic operon had more radioactivity in the aqueous phase. When this aqueous phase was resolved with

HPLC we observed several peaks, despite not expressing the $PI(4,5)P_2$ lipase PLC (Figure 3.6C). One difference between the operon and the Duet system was the plasmid backbone used to express the genes, importantly the antibiotic selection markers. The pET24 based operon utilized Kanamycin for selection, and the Duet vectors used to express this same combination of genes, pET duet and pACYC duet, utilize Ampicillin and Chloramphenicol respectively. Importantly, Chloramphenicol has been previously identified as an inhibitor of bacterial esterases, including lipases (SMITH, WORREL and SWANSON, 1949), so we hypothesized that the peaks observed in the aqueous phase of the bacteria expressing the $PI(4,5)_2$ operon may be the result of activity of bacterial lipases acting on $PI(4,5)P_2$. In order to test this, we transformed the bacterial $PI(4,5)P_2$ operon into BL21 Rosetta cells. These commercially available cells contain a plasmid encoding tRNAs that are rare in *E. coli*, and that importantly also encodes a Chloramphenicol resistance gene. When the bacterial operon was expressed in this BL21 Rosetta background with both Kanamycin and Chloramphenicol, the anomalous peaks in the aqueous phase were greatly reduced (Figure 3.6D). This result was also confirmed utilizing an empty pACYC vector which also contains a Chloramphenicol resistance gene. We found that utilization of a synthetic operon to express the genes responsible for $PI(4,5)P_2$ production is a useful alternative to utilization of the Duet vector system, however co-expression with a Chloramphenicol selectable plasmid may be preferred to inhibit any activity of endogenous bacterial lipases that may be acting on inositides.

### 3.3.5 Expression of an Inositol Importer

Another problem that has been encountered with the expression of the Inositide pathway in bacteria is the generally poor incorporation of the radiolabeled inositol into the media.

Generally, 95% or more of the material was incorporated into the cells. This is not surprising since *E. coli* do not possess a transporter capable of importing myo inositol into the cell. While this did not prevent the synthetic pathway from functioning, it does limit the total yield of inositide formation, and proves a limit for trying to investigate lower abundance products. In order to address this, we sought to include an inositol importer into the system. While *E. coli* do not possess an endogenous myo inositol metabolic pathway, there are other bacteria that do utilize myo inositol, either as a membrane lipid such as *Mycobacterium tuberculosis* and their GPI anchor pathway, or as a carbon source in the case of *B. subtilis*. While some bacteria like *Mycobacterium tuberculosis* synthesize their inositol from glucose-6-phosphate, *B. subtilis* utilize inositol from the environment. In fact, *B. subtilis* possess an entire inositol metabolic operon known as the Iol operon, and express three different transporters for the import of various isomers of Inositol. Of these YfiG and IolF have a preference for other isoforms of inositol, and IolT has a specificity for myo inositol. These transporters were all characterized in *E. coli*, and IolT expression was shown to increase the intercellular levels of Myo inositol by 300 fold (Bettaney *et al.*, 2013)(Figure 3.7A). In order to improve the expression of inositides in *E. coli* we constructed a vector for the expression of IolT from *B. subtilis* as described by(Bettaney *et al.*, 2013). When this vector was co expressed along with the Lipid phosphoinositide operon the total level of radioactivity incorporated into the cell was increased dramatically (Figure 3.7B), and the overall profile of the phosphoinositide was similar to that of bacteria without the

importer except at generally higher levels (Figure 3.7C) (note that while there is much greater incorporation of inositol in Figure 3.7B, equal CPMs are loaded in Figure 3.7C). This improvement to the expression system should allow for greater utility of the system.

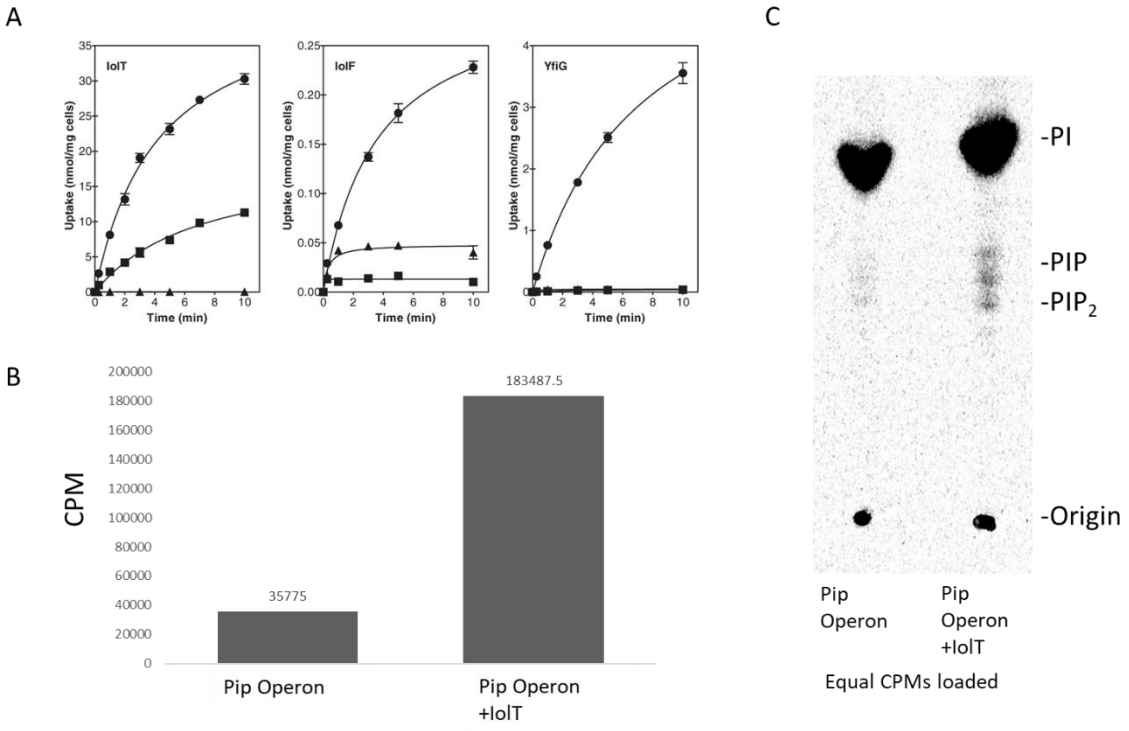


Figure 3.7 Incorporation of an inositol transporter into the synthetic inositide pathway. IolT from B. subtilis increases the efficiency of the bacterial inositide operon. A) From Bettaney et al., 2013. IolT, IolF, and YfiG all increase the incorporation of inositol in bacteria, with IolT expression causing the highest increase. B) Co expression of IolT with the bacterial inositide operon increases the ammount of $^{3}H$ Myo inositol incorporated into the lipid phase of bacterial extracts. C) Co expression of IolT with the bacterial inositide operon increases the production of PIP and $PIP_2$ when compared to the operon alone. Panel A, Copyright 2013 Taylor and Francis

### 3.3.6 Comparison of PI Kinase Activities

One of the problems encountered in the operon-based approach is that the majority of the product is the accumulation of PI, and a poor conversion of the PI to more highly phosphorylated Phosphoinositides. This was similar to the expression observed utilizing multiple vectors. In

that case the amount of PI phosphorylated by Pik1 was much lower that the amount of PI converted to PI3P by Vps34. This may partially be due to the fact that the Vps34 utilized was a truncated form that contained only the catalytic and helical domains (HELCAT) and did not contain the regulatory domains or adaptor domains. In addition to generally having lower activity, another problem may be that the bovine gene is very large, and not a lot of the protein is being produced. These two factors combined are causing a bottleneck where very little PI is being converted to PI4P, which provides very little substrate for the other enzymes in the pathway. In order to test this hypothesis, we cloned each of the PI4 kinases from yeast Pik1, Stt4 and Lsb6 into bacterial expression vectors and co expressed each of them alongside the existing operon containing the bovine Pik1. Expression of the *S. cerevisiae* Pik1 increased the production of $PI(4,5)P_2$, consistent with greater PI4P, that was then converted by Mss4 to $PI(4,5)P_2$ (Figure

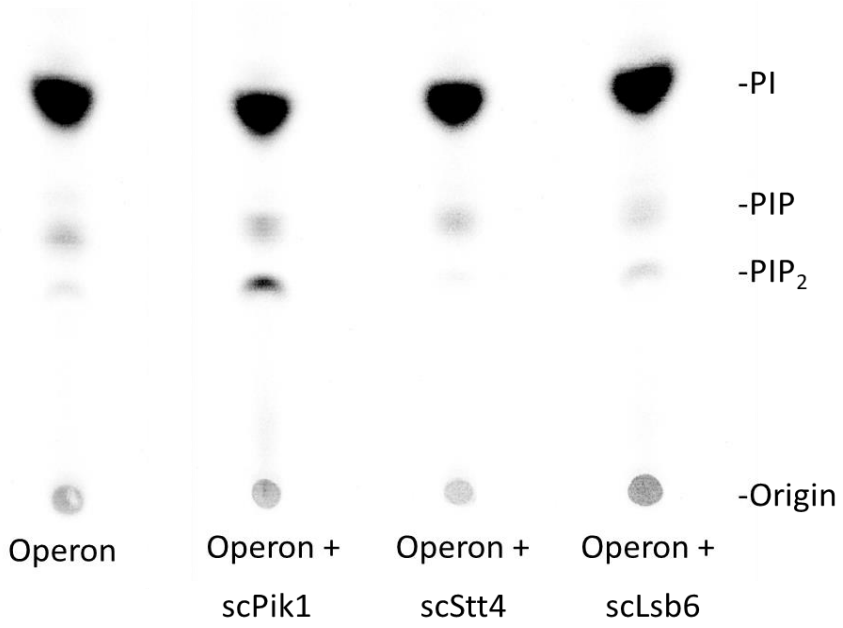


Figure 3.8 Co expression of *S. cerevisiae* PI 4 kinases along with the PIP operon. Expression of the bacterial inositide operon alone (Lane 1), with scPik1 (Lane 2), with scStt4 (Lane 3) or with scLsb6 (Lane 4). When scPik1 is co expressed along with the bacterial inositol operon increases the formation of $PIP_2$, indicating that scPik1 better converts PI to PI4P, which then is converted to $PIP_2$ by Mss4.

3.8 lane 2 vs lane 1). Expression of Stt4 and Lsb6 had little to no improvement of the system, consistent with previous reports of their lower activity (Figure 3.8 lanes 3 and 4 vs lanes 1 and 2).

## 3.4 Discussion

Overall, these data indicate that in addition being able to recapitulate lipid inositide signaling, *E. coli* are capable of reconstituting the soluble inositide signaling pathway as well. Due to the important roles that inositides play in many biological processes, better understanding of the enzymatic activities controlling their production and breakdown is essential to delineating their roles in normal cellular function and in disease states. However, because eukaryotic systems are exceedingly complex, utilization of a simpler cell-based system can provide a unique tool to elucidate precise enzymatic functions that may be otherwise difficult to ascertain. While *in vitro* analysis is useful for confirmatory experiments, a cell-based system enables reductionist-based approaches in a biological context. To this end, we provide evidence that inositide signaling can be successfully reconstituted in *E. coli*, offering a cell-based platform devoid of endogenous inositide signaling to enable characterization of inositide kinases, phosphatases, and regulatory proteins.

While this system was able to reconstitute the canonical inositide synthesis pathway, we

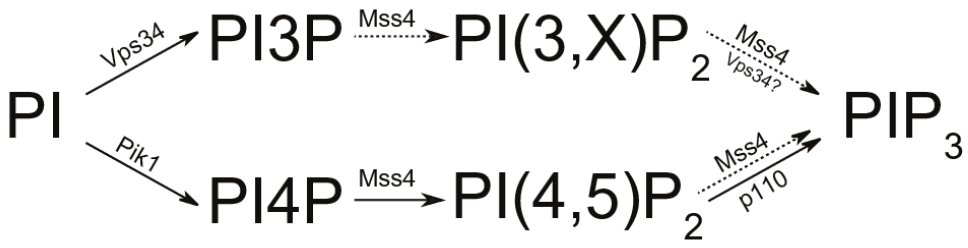


Figure 3.9. Mss4 can act to produce $PIP_3$ in two separate pathways. Mss4 acts as promiscuous multikinase in two separate pathways for the production of $PIP_3$

also observed a heretofore unappreciated route to $PIP_3$ formation with expression of only Pis1, Pik1, and Mss4 (Figure 3.2C and Figure 3.3H). While Mss4 has been previously annotated as a canonical PI(4)P 5-Kinase, to our knowledge, Mss4 has not been described as a PI(4)P 3-kinase or as a $PI(4,5)P_2$ 3-kinase as our data suggests. However, an alternative explanation of these data is the presence of an endogenous bacterial lipid kinase. We do not favor this interpretation as, to our knowledge, there is no putative bacterial inositide kinase based on sequence homology, although such a finding would be of interest. Here we describe two distinct pathways for the conversion of PI to $PIP_3$, and both are dependent on multiple kinase activities of Mss4 (Figure 3.9). Description of a previously unrecognized PIP multikinase activity of Mss4 would not have been possible without an *in vivo* system such as the one described here.

Although the primary pathway to produce $PIP_3$ in mammalian cells is through PI(4)P and $PI(4,5)P_2$, there is also evidence that a PI(3)P dependent pathway to $PIP_3$ is present in both mammalian cells (Halstead *et al.*, 2001) and fission yeast (Mitra *et al.*, 2004) through the *Schizosaccharomyces pombe* Vps34 and Mss4 homologues. In order to demonstrate this alternative pathway to $PIP_3$, we expressed a truncated form of Vps34 along with Pis1 and were able to produce PI(3)P (Figure 3.2B). Further, with the addition of Mss4 we observed both $PI(3,X)P_2$, and $PIP_3$ (Figure 3.2C and Figure 3.3G). These data recapitulate this proposed alternate pathway to produce $PIP_3$ with robust conversion of PI to $PIP_3$ via PI(3)P and $PI(3,X)P_2$ intermediates (Figure 3.9). While it is clear that the formation of the $PI(3,X)P_2$ is catalyzed by Mss4, since Vps34 alone only produces PI(3)P (Figure 3.2A, 2B and Figure 3.3E), the conversion of $PI(3,X)P_2$ to $PIP_3$ could be catalyzed by either Mss4 or Vps34. Even though this PI(3)P-based pathway to $PIP_3$ in this *E. coli* system was very robust, it is a secondary pathway in mammalian cells, a minor pathway in fission yeast only detectible after deletion of phosphatases,

and interestingly, has yet to be observed in budding yeast. We believe this demonstrates the utility of our system to expose the activity of the enzymes free from alterations in cell compartmentalization, phosphatases, or other regulatory factors.

Reconstitution of the inositide pathway in *E. coli* revealed the formation of $PIP_3$, an inositide not observed in *S. cerevisiae*, despite utilizing only *S. cerevisiae* enzymes (Figure 3.2C lane 1, Figure 3.3G). One possibility for this observation is the difference in spatiotemporal distribution of enzymes and their lipid substrates. A prokaryotic system enables characterization of inositide enzyme activities in a largely "noise-less" experimental context, where the kinases, phosphatases and regulatory proteins may have access to lipids without consideration for subcellular localization. Additionally, in eukaryotes the inositide pathway is highly regulated by phosphatases, while in this bacterial system there are no endogenous inositide phosphatases. In *S. cerevisiae*, a pathway to generate $PIP_3$ may exist, but is suppressed by phosphatases that act on either $PIP_3$ or the $PI(3,X)P_2$ intermediate.

While *E. coli* do not contain endogenous inositide signaling, there are still some potential caveats worth discussing. First, *E. coli* do encode an inositol 1-phosphatase, CysQ, that has homology to the mammalian $I(1,4)P_2$ 1-phosphatases bisphosphate nucleotidase (Bpnt1) and inositol polyphosphate 1-phosphatase (INPP1). *In vitro* analysis indicates that CysQ's basal activity for $I(1,4)P_2$ is low, and its *in vivo* activity in this bacterial system is minimal, as we still observe appreciable signal where $I(1,4)P_2$ elutes on HPLC (Figure 3.4C). Second, *E. Coli* contain bacterial phospholipases, and we observe the formation of small amounts of lyso-PI in some of our bacterial extracts examined by TLC (Figure 3.2A). However, combining TLC analysis of inositide lipids with HPLC analysis of deacylated glycerol-inositide phosphates, we are able to resolve lyso-PI from other species (Figure 3.3). We also observed that some phosphatidyl

inositol phosphates seem to be cleaved by bacterial lipases, and further that this cleavage could be inhibited by chloramphenicol (Figure 3.6 C and D) consistent with previous observations about chloramphenicol (SMITH, WORREL and SWANSON, 1949). Finally, the presence of bacterial phytase may catalyze the formation of additional IP species, although its overall effects appear to be small as the levels of $IP_6$ produced are much greater than the levels of possible phytase products (Figure 3.5G). However, utilizing this system as a discovery tool for IP kinase/phosphatase activities may necessitate the inhibition of this phytase either genetically or pharmacologically.

In addition to studying the process of making PIPs and IPs, this synthetic system could be used to study the proteins that interact with PIPs and IPs. Many proteins have been identified as PIP-binding partners (Blind *et al.*, 2014; Sablin *et al.*, 2015; Sun and MacKinnon, 2017; Delgado-Ramírez *et al.*, 2018), and this system could be used to co-express and study these proteins along with their endogenous binding lipids *in vivo*, without the complex environment of the eukaryotic cell. In the case of IPs, a number of proteins have been shown to rely on $IP_6$ as a structural cofactor (Macbeth, Schubert, VanDemark, Lingam, Hill and Bass, 2005; Sheard *et al.*, 2010; Ouyang *et al.*, 2016b) as it was discussed in Chapter 2. For these proteins, expression of protein for structural studies was not possible using *E. coli* since they lack $IP_6$, a key factor for folding. This synthetic system may allow for structural studies of proteins that utilize IPs as cofactors to be conducted in *E. coli* rather than other expression systems.

While here we express portions of the synthetic inositide pathway as an operon-based plasmid, future iterations of this system can be envisioned as stably expressing the entire pathway in a synthetic operon integrated into *E. coli* genome. Initially we observed relatively inefficient radiolabeling with inositol, frequently having greater than 95% of the $^3$H-inositol left

in the medium. Addition of an inositol transporter IolT greatly increased the yield. Our data are a proof of concept that *E. coli* can be modified to express both the PIP and IP pathways, and can provide a useful tool for discovery and characterization of key inositide regulatory processes.

# Chapter 4 Characterization of the $IP_6$ Binding Protein Clu1

## 4.1 Introduction

In addition to their classically studied role as second messengers, there is emerging evidence that IPs may play additional non-signaling roles. Our lab has developed a kinase assay based screen to identify proteins utilizing Inositol hexakisphosphate ($IP_6$) as a structural cofactor (Pham, 2012). This screen identified two proteins NatA, an N-Terminal Acetylase, and Clu1 as proteins that utilize $IP_6$ as a structural cofactor (Pham, 2012). Several other proteins that bind IPs as structural cofactors have been identified incidentally when their crystal structures were solved with $IP_6$ buried inside solvent inaccessible pockets within the protein (Macbeth et al, 2005) (Sheard et al, 2010) (Ouang et al, 2016). In addition, our laboratory has previously solved the crystal structure of NatA (Neubauer, 2012), as well as other $IP_6$ binding Nat complex assemblies (PDB 4XNH). Conversely, very little was previously known about the structure and function of Clu1. This dissertation describes the work to characterize Clu1 and the role of $IP_6$ in its function, determine the structure of Clu1 and describe the role that $IP_6$ plays in its structure, additionally further describing the non-signaling role of $IP_6$.

Clu1 was first identified in *Dictyostelium*, where deletion of the Clu1 homologue CluA resulted in the alteration of mitochondrial morphology, a phenotype where the mitochondria were clustered at the center of the cell (Zhu *et al.*, 1997). In budding yeast Clu1 was initially identified as a translation initiation factor named Tif31 (Vornlocher *et al.*, 1999), however this role in translation initiation was later ruled out (Hinnebusch, 2006), and it was renamed Clu1 since deletion of Clu1 in *S. cerevisiae* resulted in a similar clustered mitochondrial phenotype to *Dictyostelium* (Fields, Conrad and Clarke, 1998). It was also shown that heterologous

expression of yeast Clu1 in CluA null *Dictyostelium* rescued the altered mitochondrial morphology (Fields, Conrad and Clarke, 1998).

Much of what is known about Clu1 function comes from characterization of the *Drosophila melanogaster* homologue of Clu1 called Clueless (Cox and Spradling, 2009). The clustered mitochondrial phenotype observed in yeast and *Dictyostelium* was also observed in *Drosophila* with disrupted Clueless expression (Cox and Spradling, 2009). It was also shown in *Drosophila* that Clueless interacts genetically with the most commonly mutated gene in inherited Parkinson's disease, Parkin (Cox and Spradling, 2009). It was later also shown that Clueless interacts physically with Parkin as well as the serine threonine kinase Pink1 (Sen *et al.*, 2015). Collectively, these studies implicated Clu1 in recognition of and response to damaged mitochondria. This idea was further enhanced by the observation that Clueless works to clear mitochondrial damage by interacting with mitochondrial quality control pathway proteins Mitochondrial Assembly Regulatory Factor (Marf) and Valosin-Containing Protein (VCP) (Wang, Clark and Geisbrecht, 2016). These damage response pathways initiate the destruction and recycling of damaged mitochondria and damaged mitochondrial proteins in response to oxidative damage.

In *Arabidopsis thaliana*, a Clu1 homologue was identified and named friendly mitochondria (Fmt), and when deleted the mitochondria were clustered, like in flies, *Dictyostelium* and budding yeast (Logan, Scott and Tobin, 2003). *Arabidopsis* also have another Clu1/Fmt homologue, Reduced Chloroplast coverage (Rec1) (Larkin *et al.*, 2016). Deletion of Fmt1 or Rec1 result in a variety of chloroplast defects including abnormal localization, reduced chloroplast size, inability of chloroplast to localize in response to stimuli as well as clustering of mitochondria (Larkin *et al.*, 2016). This indicates that in plants Clu1 homologues may play a

larger role in regulating the metabolic state, as well as in the localization of subcellular energetic organelles.

In addition to characterization in *Drosophila*, *Dictyostelium* and plants, Clu1 has also been characterized in mammalian cells, where its homologue is called Cluh (Gao *et al.*, 2014). Disruption of Cluh expression in mammalian cells results in the clustered mitochondria phenotype observed in other species. It was observed in mammalian cells that Cluh was an RNA binding protein that was associated with mRNAs encoded by nuclear genes for protein targeted to the mitochondria (Gao *et al.*, 2014). This led to the hypothesis that Cluh functions to help these targeted mRNAs get to the mitochondria surface, possibly to endoplasmic reticulum (ER)-mitochondrial contact sites where they are translated, and the resulting peptides are imported into the mitochondria. This was bolstered by the observation that Cluh coupled the energetic state of the cell with the metabolic state (Wakim *et al.*, 2017). It was also observed that Cluh is required for infectivity and transport of viral RNA by the influenza virus, indicating that the RNA binding function of Cluh has been hijacked by viruses (Ando *et al.*, 2016; Švančarová and Betáková, 2018). A Cluh knockout mouse was generated and the mice were completely normal until birth, at which point the mice died within a few hours (Schatton *et al.*, 2017). It was established that the reason the mice died was that they were unable undergo oxidative phosphorylation after birth.

Overall these studies describe Clu1 as playing a role in recognizing and responding to oxidative damage to mitochondria and mitochondrial proteins, though other than being associated with proteins involved in this process, the mechanism of action for Clu1 in these processes has yet to be described. Clu1 also interacts with mRNAs important for oxidative phosphorylation, and this RNA binding can be exploited by viruses. When Clu1 is deleted the

mitochondria are mislocalized into clusters, which in metazoans seems to be required for life. In single cell organisms, while required for mitochondrial distribution it seems to be largely dispensable for mitochondrial function. Evidence from plants indicate that it plays a role in the function and localization of energetic organelles including both mitochondria and chloroplasts, demonstrating a larger role in the management of cellular metabolism at the organelle level.

A previous student in the York lab also identified Clu1 as a protein that binds to $IP_6$ as a structural cofactor (Pham, 2012). By further characterizing this protein, its RNA binding, and the role of $IP_6$ in its function we can better understand the role that is plays in response to mitochondrial damage like in Parkinson's disease, and its role in regulating response to metabolic changes, where it has been shown to be required for multicellular life. This chapter describes the characterization of Clu1, its RNA binding function, its $IP_6$ binding, and the role that both of these play in the overall structure and function of Clu1 within yeast cells. The next chapter describes our structural characterization of Clu1 and the importance of $IP_6$ binding to its structure.

## 4.2 Methods

### 4.2.1 Cloning

Full length Clu1 was amplified from *Saccharomyces cerevisiae* W303 with PCR and cloned into pKLMF-EK vector (NEB) downstream of maltose binding protein (MBP). An additional cleavage site for Tobacco Etch Virus (TEV) protease was introduced between MBP and Clu1 sequence that resulted in the following sequence between MBP and Clu1: LEENLYFQG**SEKKEE** (TEV site underlined, Clu1 sequence in bold). The plasmid fragment containing MBP-Clu1 sequence was subcloned into *S. cerevisiae* p415 vector under either TEF

or GPD constitutive promotors (Mumberg, Müller and Funk, 1995). This same method was used to generate C-terminally truncated Clu1 constructs as well.

### 4.2.2 Stable Knockout of Clu1 C-terminus

Stable knockout of Clu1 C-terminus was conducted using a homology-based recombination strategy. Briefly primers were designed to amplify the Trp1 gene with 75 bp extensions with homology to the sequences upstream of Clu1 at its C-terminus, specifically with the downstream primer beginning 50 bp upstream of the Clu1 C-terminus introducing a premature stop codon to produce a truncated form of Clu1. After PCR amplification, the construct transformed into W303 yeast, and colonies that integrated the construct were selected on CSM -Trp media and verified with PCR.

### 4.2.3 Growth of Yeast

For all yeast growth standard selection (CSM) drop out media and rich media (YPD) were used except when indicated. Expression vectors were transformed into yeast using standard lithium acetate transformation (Ito *et al.*, 1983). Briefly, 1 ml cultures of the yeast strain to be transformed were grown overnight in YPD or appropriate growth media. Cells were spun down and media was removed. Approximately 1ug of desired plasmid and 10ug of boiled herring sperm DNA were added to the cell pellet and the pellet was resuspended with a Plate Mix consisting of 40% PEG4000, 100mM LiOAc, 10mM Tris pH 7.5 and 1mM EDTA pH 8.0, and incubated with rotation overnight. The next day 40µl of DMSO was added and the cells were heat shocked at 42°C for 15min, and then the cells were spun down and the mix removed. Cells were then resuspended in water and then plated on the appropriate selection plates.

### 4.2.4 Protein Purification

For purification of Clu1FL and Clu1ΔC50 an overnight starter culture of W303 clu1Δ yeast were transformed with the appropriate vector for expression of the fusion protein and was grown overnight in CSM -Leu media at 30°C. After this YPD media was inoculated with 1ml of the overnight starter per liter of culture and grown for 24 hours at 30°C. Cells were then harvested and washed with water before freezing at -80°C and stored there until use. For purification, cells were thawed on ice, and then resuspended 5:1 with 50 mM Tris pH 7.5 200 mM NaCl 1 mM EDTA with Protease inhibitors (Roche tablets). The cells were then lysed by passing 3x through an EmulsiFlex homogenizer (Avestin) at >25000 PSI. This crude lysate was then cleared by centrifugation at 20000 xg for 30 min at 4°C. To the clear lysate Amylose Fast Flow resin (NEB) was added 100 µl of 50/50 slurry in lysis buffer per gram of cell pellet. The lysate was incubated with the slurry for 1 hour with rotation at 4°C. The mixture was then added to a chromatography column and the unbound lysate was allowed to flow through. The resin was washed with 20 column volumes of Lysis buffer, and the Clu1 was eluted with Lysis buffer + 10 mM maltose. The MBP tag was then removed by digestion with TEV protease (Either with RNaseA for removing RNA from FL Clu1 or without). The protein was then further purified using a Superdex 200 Increase 10/30 size exclusion column. The resulting protein was concentrated with a Spin Concentration device (Millipore).

### 4.2.5 RNA Purification and Analysis

To purify RNAs bound to Clu1, Full length Clu1 protein was purified as described above. After elution from the Amylose resin, the RNA was immediately extracted with

Phenol:Chloroform:Isoamyl Alcohol 25:24:1 and precipitated with Ethanol using Glycogen as a carrier. For nuclease digestion the RNA was resuspended in TE buffer and treated with Nuclease as indicated. Gel analysis of RNA was conducted with 12% acrylamide Urea, or 1.2% Agarose TBE gels, and RNA was visualized with SYBR gold. For RNA sequencing RNA pellets from ethanol precipitation were sent to VANTAGE at Vanderbilt, and the resulting sequences were processed utilizing TopHat2, Bowtie (Kim *et al.*, 2013) and DEseq (Love, Huber and Anders, 2014).

### 4.2.6 Analytical Gel Filtration Chromatography

Analytical Gel filtration was conducted using a Superdex 200 Increase 10/30 column. A flowrate of 0.2 ml/min at 4°C was used. Elution volumes of the column were calibrated with molecular weight standards as described by the manufacturer. Elution was monitored with in line absorbance readings at 280nm to measure amino acid side chains, and 254 nm to measure nucleic acids. Additionally, 215 nm was also utilized for low concentration samples to measure the peptide bond.

### 4.2.7 IP Kinase Assays

For IP kinase assays protein samples were added, either boiled for 5 minutes, or un-boiled with Vip, or Ipk1 and $^{32}$P-γ-ATP in 50mM Bis-Tris pH 6.0 10mM $MgCl_2$. The reaction was incubated at 37°C for 1hr. The reactions were stopped by addition of 2.1M HCl, and then the entire reaction was spotted onto PEI-F Cellulose. The reactions were separated by TLC with 2.1M HCl for Ipk1 assays or 2.25M HCl for Vip assays with $KH_2PO_4$ 1.09M and $K_2HPO_4$ 0.72M tank buffer until the solvent front reached about 1 cm from the top of the plate. The plates were

allowed to dry, and then visualized by phosphorimaging with a Typhoon 9000 imager.

### 4.2.8 Dynamic Light Scattering

Dynamic light scattering analysis of Clu1 was conducted using Wyatt DynaPro NanoStar instrument. Measurements were conducted at room temperature using 10 µl of 1 mg/ml Clu1 treated with RNaseA and Clu1ΔC50, in NanoStar disposable cuvettes.

### 4.2.9 Analytical Ultracentrifugation

Clu1ΔC50 protein purified as described above was analyzed using an Optima XLI ultracentrifuge (Beckman Coulter, Brea, CA) using with a four-hole An-60 Ti rotor at 25,000 RPM overnight at 4°C. The protein was loaded into dual sector 1.2cm pathlength cells with sapphire windows. Data was analyzed using Sedfit (Schuck, 2000).

### 4.2.10 Negative Stain Electron Microscopy

For negative stain electron microscopy analysis of Clu1ΔC50, 4 µl of 0.01 mg/ml of protein was applied to carbon coated copper grids, for 60 seconds. Excess protein was removed by blotting with filter paper. The grids were then washed 2x with water, and then once with Uranyl formate solution (0.7% w/v), and then stained with Uranyl formate solution for 90 seconds. Excess stain was removed by blotting and aspiration. Negative stain data was collected at 200kV with a FEI Tecani TF20 with a CCD camera. Images were collected at 62000x and a pixel size of 1.757Å. Negative stain data was analyzed with the EMAN2 software package (Tang et al., 2007). Particles were selected with semiautomated picking, and 2D class averages were generated.

### 4.2.11 Native Mass Spectrometry

Flor native mass spectrometry analysis 50mM Clu1ΔC50 protein was buffer exchanged into 100mM ammonium phosphate and analyzed with FT-RCR Mass Spectrometry in collaboration with Boone Prentice at the Vanderbilt University Mass Spectrometry Research Center.

### 4.2.12 Microscopy

Microscopy was conducted using a Nikon Eclipse TE200E and a 100x objective lens. Yeast were prepared for microscopy by transforming with a mtGFP plasmid as described in (Westermann and Neupert, 2000). Colonies were then selected and grown overnight at 30°C in appropriate selective media. The yeast were then washed with and transferred to YP+2%Glycerol media, and allowed to continue growing at 30°C for 3hr. At this point 5µl of cells in media were added to 5µl of YPGlycerol in low melting point agarose on a microscope slide and a cover slip was applied. The slide was allowed to cool and settle for about 10 minutes before imaging. Images were acquired with MetaMorph, and were processed using the Fiji distribution of ImageJ (Schindelin *et al.*, 2012).

### 4.2.13 Spot Assays

Plates for spot assays were prepared as follows. Yeast Extract and Peptone agar was prepared and autoclaved. When this mix had cooled to approx. 65°C 2% (v/v final concentration) of filter sterilized Ethanol, Lactose pH 5.5, or glycerol were added and mixed before pouring plates. Overnight cultures of the appropriate yeast strains were diluted to an OD600 of 1.0 in milli Q

H2O, and then serially diluted 1:10 across 8 conditions. Then 5µl of each dilution was plated on select conditions and grown at the indicated temperature.

## 4.3 Results

### 4.3.1 Clu1 RNA Binding

In addition to binding to $IP_6$, Clu1 in humans was identified as binding to RNAs, specifically mRNAs for nuclear encoded mitochondrial targeted genes. We sought to

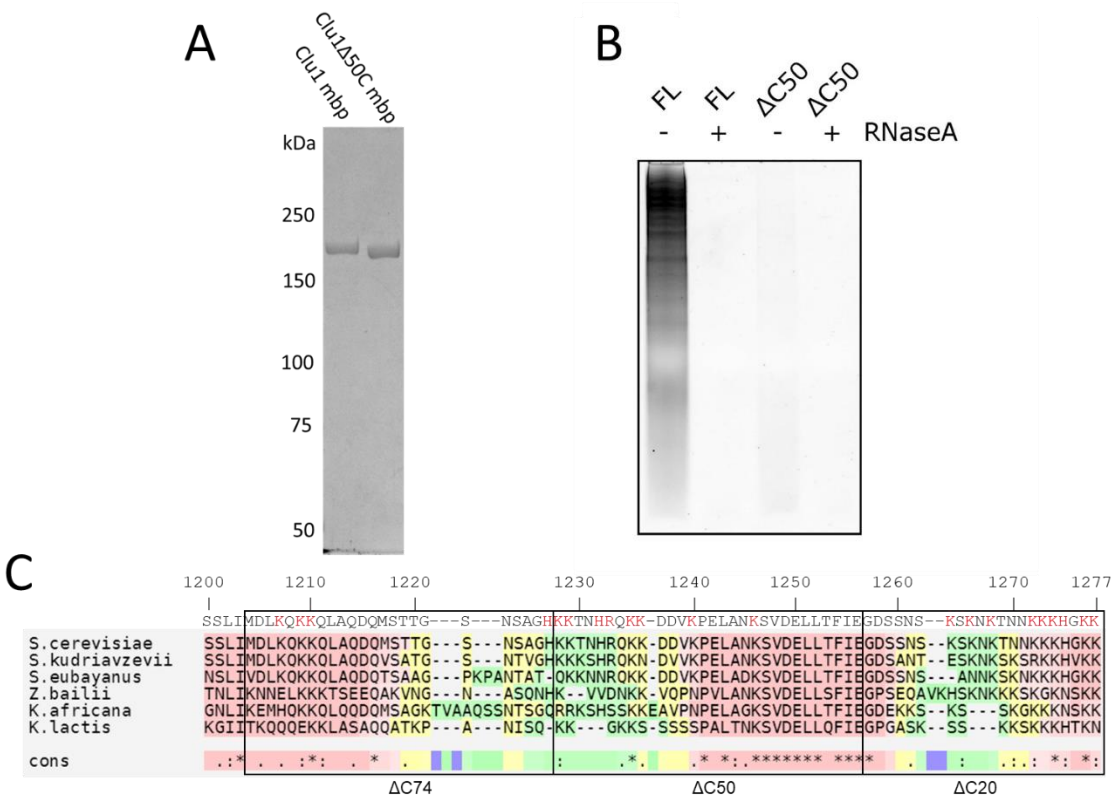


Figure 4.1 Truncation of Clu1 eliminates RNA binding. A) Purified Clu1 FL and Clu1ΔC50 with MBP fusion protein separated with 6% Acrylamide SDS Gel and visualized with Coomassie Blue. B) Purified Clu1 was treated with RNaseA, extracted with Phenol Chloroform and Ethanol precipitation, separated on 12% acrylamide denaturing Urea gel and visualized with SYBR gold. C) Fungal alignment of C- termini of Clu1 homologues highlighting clusters of positive residues (top row) conserved sequences (middle) and deletions tested (boxes and bottom row).

characterize the RNA binding function of Clu1, identify the portion of the protein responsible for RNA binding, and determine if RNA binding and IP binding were dependent on each other. In the course of preparing expression constructs for crystallography of Clu1, a variety of truncation mutants were generated, particularly truncations of the C-terminus, which contained many lysine residues and was predicted to be unstructured. It was observed that truncations of the C-Terminus of Clu1 by 20 amino acids, 50 amino acids and 74 amino acids decreased the co purification with RNA, and that a deletion of the C-Terminal 50 amino acids is sufficient to almost completely remove copurification of RNA with the protein (Figure 4.1A, 1B). An alignment of this C-terminal region of Clu1 homologues from several fungal species shows that the C-terminal 50 amino acids includes a conserved patch of amino acids, as well as several less conserved patches of positively charged amino acids, including many clusters of Lysine residues (Figure 4.1C). It is possible that either the positively charged patches, or the conserved sequence, or both may be responsible for Clu1's interaction with RNAs. It was also observed that Clu1 purified from yeast co purifies with a significant portion of RNA based on absorbance of the sample at A254 compared to absorbance at A280 when the protein is analyzed with an s200 Size exclusion column (Figure 4.2A and B solid line), and the protein elutes in the void volume of the column indicating that Clu1 bound to RNA represents a large complex eluting near or in void volume of the column. If the sample is digested with RNAseA prior to separation, the absorbance at 254 nm decreases dramatically (Figure 4.2B dashed line), and the molecular weight of the Clu1 complex in the absence of RNA is much lower (Figure 4.2A dashed line). If the Clu1ΔC50 construct is purified, it elutes at the same time as the RNAseA treated FL protein (Figure 4.2A dotted line) and co purified with almost no RNA based on absorbance at 254nm (Figure 4.2B). Both the RNaseA treated FL protein and the Clu1ΔC50

protein still elute earlier than the anticipated molecular weight of a Clu1 monomer, indicating that Clu1 exists in some multimeric form.

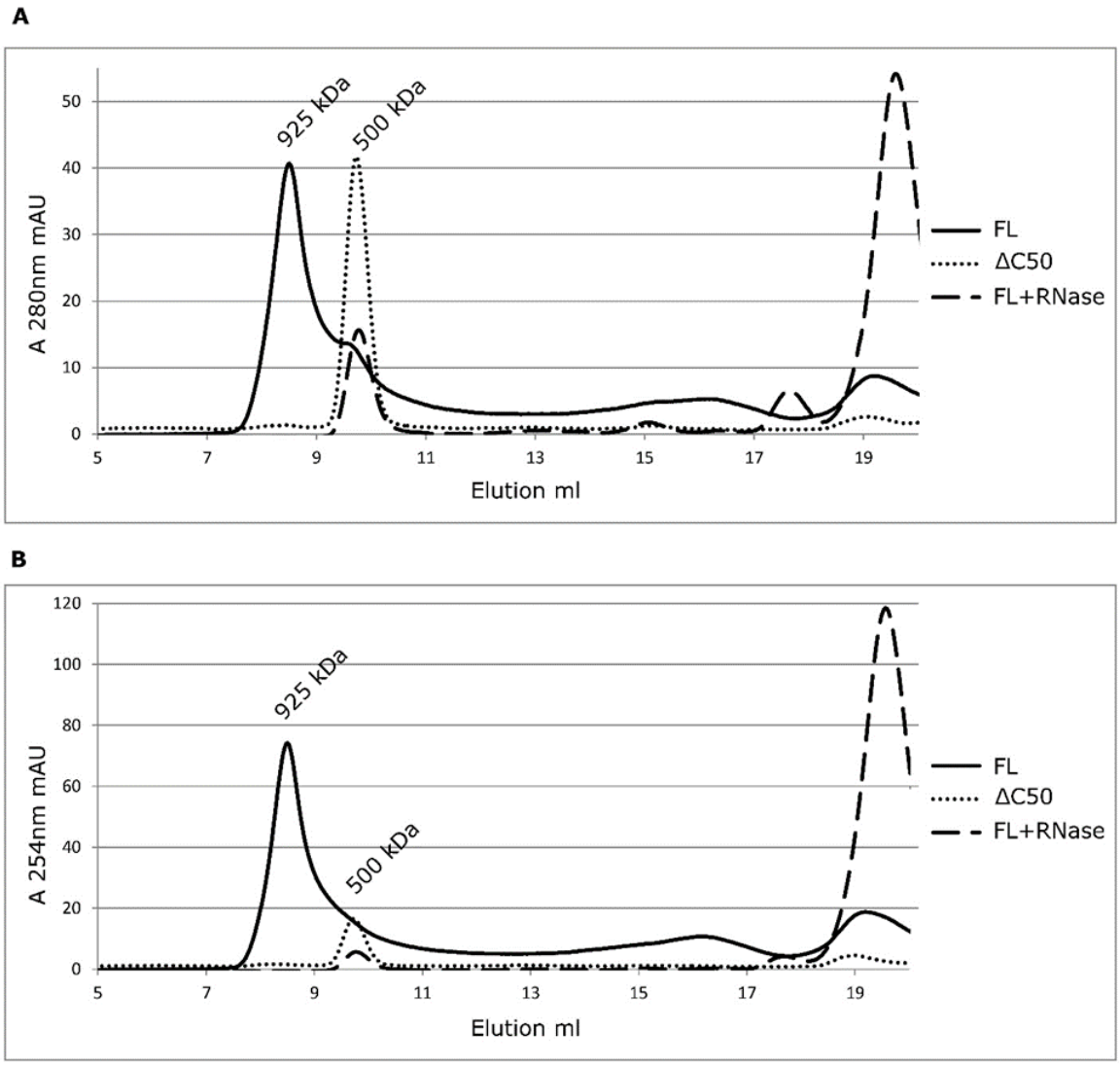


Figure 4.2. Size exclusion chromatography of full length and ΔC50 Clu1. A) Absorbance at 280nm showing FL Clu1 elutes at a high molecular weight near the void volume, while RNAse treated and ΔC50 Clu1 elute at a about 500kDa. B) Absorbance at 254nm showing a large peak of nucleic acid co elutes with FL but not ΔC50 Clu1, and upon treatment with RNase the nucleic acid peak is gone.

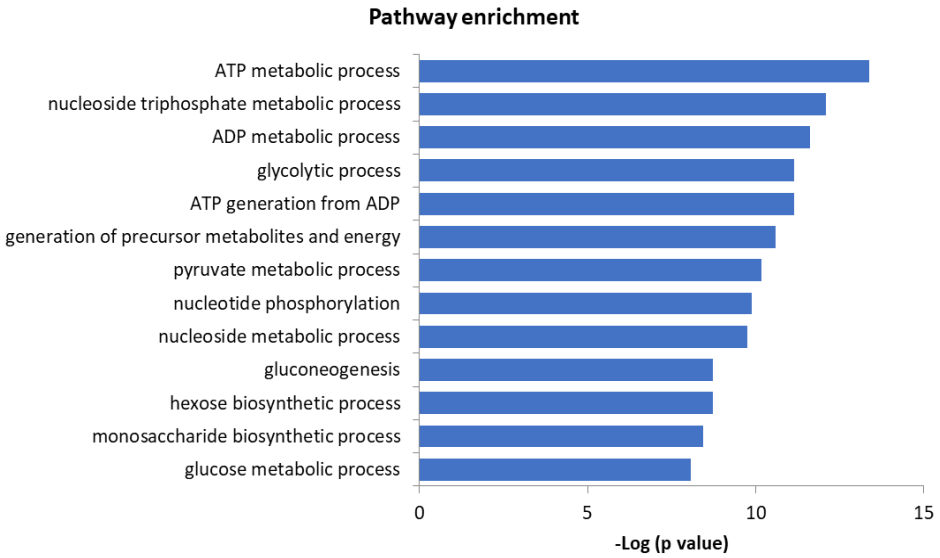


Figure 4.3 Clu1 binds to RNA associated with metabolic processes. RNA seq was conducted on RNA that co purified with Clu1, and the RNAs identified are associated with a variety of energetic pathways.

We also sought to identify the RNAs that were bound to Clu1 in yeast. To this end we isolated the RNA that co purified with Clu1 and conducted RNAseq. The isolated RNA was sequenced with VANTAGE at Vanderbilt University, and the resulting reads were processed with TopHat2, Bowtie and Cufflinks (Kim *et al.*, 2013). We identified a set of mRNAs that were enriched over total RNA using DESeq2 (Love, Huber and Anders, 2014) with a p-value $< 0.05$, and the gene ontology of these mRNAs were enriched for mitochondrial and glucose metabolic pathways with a p-value $< 0.01$. (Figure 4.3). This analysis agrees with previous work in mammalian cells that identified mRNAs for a variety of mitochondrial processes (Gao *et al.*, 2014).

To further define the RNA binding role of Clu1, we constructed an expression vector to express only the C-terminal 50 amino acids of Clu1 along with the MBP tag referred to as the C50 construct, as well as the MBP tag alone as a control (Figure 4.4A). When purified, both the full length Clu1 protein and the C50 construct co purified with RNA (Figure 4.4B), although the

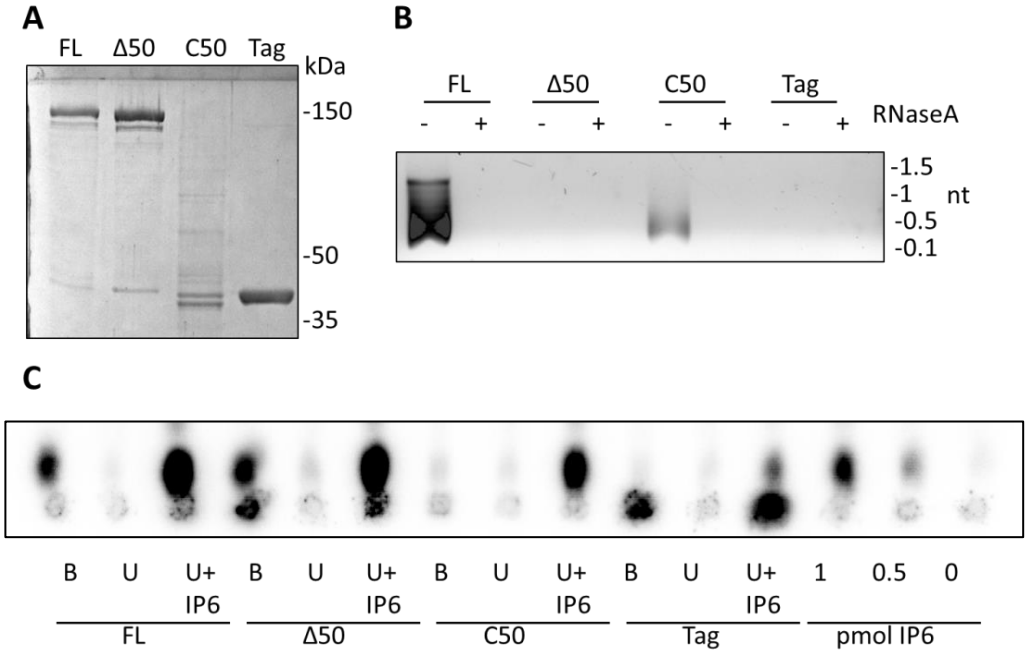


Figure 4.4 Clu1 binds to RNA with its C-terminal 50 amino acids. A) Clu1 constructs on SDS PAGE showing Full Length Clu1, Clu1 C terminal truncation, Clu1 C terminal 50 Amino Acids, and MBP tag alone. B) RNA purified from Clu1 constructs treated with RNaseA then purified with Phenol Chloroform and Ethanol precipitation, separated on 1.5% agarose TBE gel and visualized with SYBR Gold. C) Kinase assay on Clu1 constructs either Boiled (B), Unboiled (U) or Unboiled with $IP_6$ added (U+).

C50 construct copurified with less RNA than the full-length protein. The Clu1Δ50 protein and the MBP tag alone were not capable of copurifying with RNA (Figure 4.4B). After identifying that the RNA binding domain is at the C-terminus of Clu1, we wanted to investigate the role, if any, that RNA binding had on the $IP_6$ binding of Clu1, since it had been shown that the TPR domains near the C-terminus of the protein were responsible for $IP_6$ binding (Pham, 2012). In

order to test this, we conducted Vip kinase assays on Clu1 full length protein, Clu1ΔC50 protein, C50 construct and MBP tag to detect $IP_6$ bound to the protein (Figure 4.4C). This assay showed that both the full length Clu1 protein, and the Clu1ΔC50 protein were able to bind to $IP_6$ in a boil dependent manner. Neither the C50 construct nor the tag alone were capable of binding to $IP_6$ (Figure 4.4C).

### 4.3.2 Clu1 IP Binding

Next we wanted to investigate the nature of $IP_6$ binding to Clu1, and determine both the stoicheometry of Clu1 to bound $IP_6$, as well as the ability of Clu1 to bind to other IPs, and if the

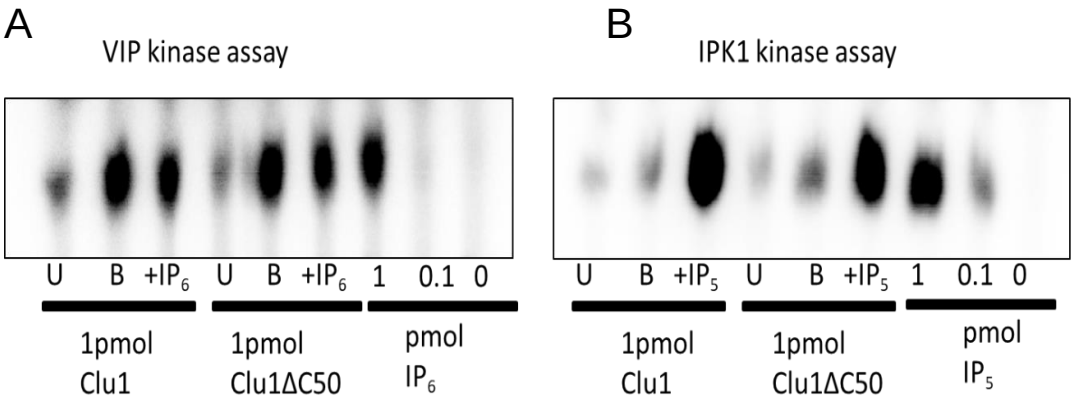


Figure 4.5 Clu1 binds to $IP_6$ and $IP_5$. Clu1 is bound to IP at a 1:1 ratio, primarily $IP_6$, but with a small population of $IP_5$. Both full length and ΔC50 Clu1 are bound to IPs, and at the same ratios.

other IPs are able to functionally replace $IP_6$ binding. When 1pmol of either full length or ΔC50 Clu1 were tested in a Vip kinsae assay 1 pmol of $IP_6$ was found to be bound (Figure 4.5A). This indicates that $IP_6$ is bound to Clu1 at a 1:1 molar ratio in both the Full length and ΔC50 protein. Previously $IP_5$, which is very similar to $IP_6$, has been shown to bind to some $IP_6$ binding proteins, so we conducted a kinase assay utilizing the $IP_5$ kinase Ipk1 to determine the extent of $IP_5$ bound to Clu1. We found that a small amount of $IP_5$ bound, but that $IP_6$ was the major IP species bound in both the full length and the ΔC50 Clu1 (Figure 4.5B). This indicates that in wild type

conditions most Clu1 is bound to $IP_6$, but does not establish if $IP_5$ is capable of fully replacing $IP_6$ for Clu1 function when $IP_6$ is not avalable.

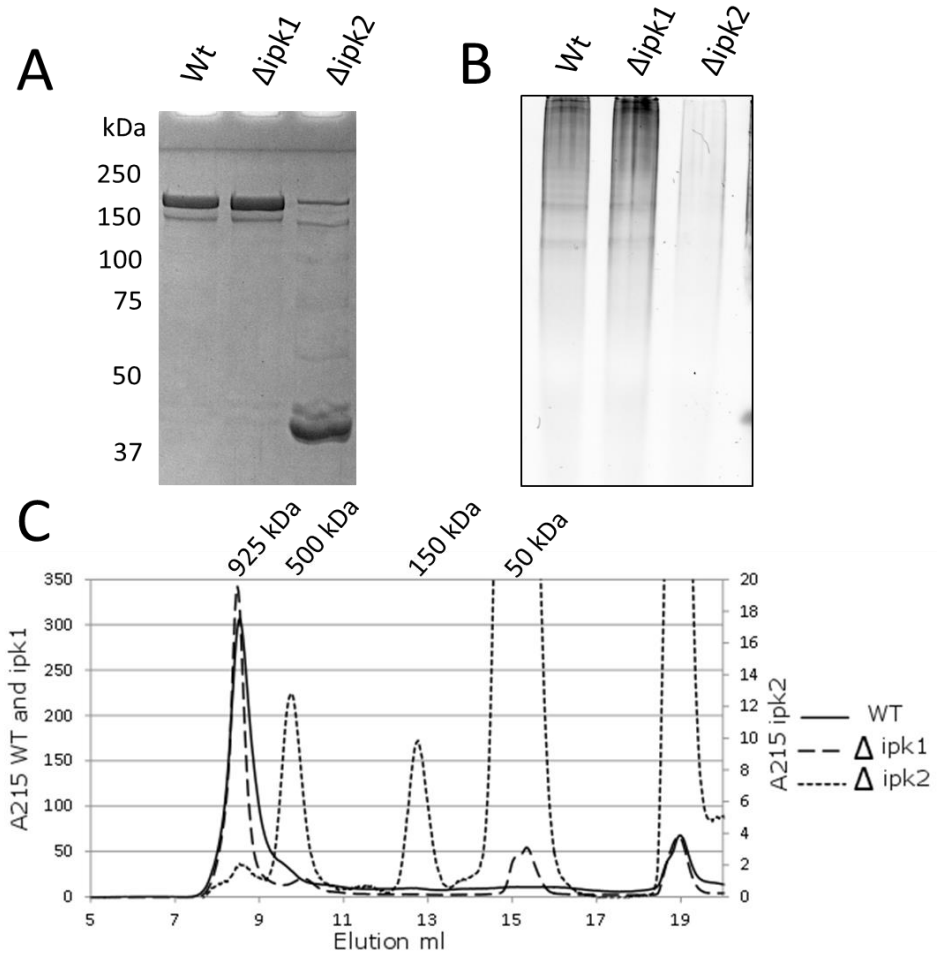


Figure 4.6 IP binding is required for Clu1 stability. Deletion of Ipk2, but not Ipk1 impacts the stability and RNA binding of Clu1. A) SDS PAGE showing degradation of Clu1 purified from Δipk2 yeast. B) RNA binding to Clu1 is decreased in Clu1 purified from Δipk2. C) Size exclusion chromatography of Clu1 shows that the size of Clu1 purified from Δipk2 yeast is altered.

We next used a genetic approach to remove $IP_6$ to determine if $IP_5$ or other IPs were capable of replacing $IP_6$. We utilzed ipk2Δ yeast, which cannot phosphorylate $IP_3$ and thus lack $IP_4$, $IP_5$ and $IP_6$. We also used ipk1Δ yeast which produce $IP_5$ but not $IP_6$. We expressed and purifed Clu1 protein from these strains and then analyzed the protein. From ipk1Δ yeast we were able to purify nearly the same ammount of protein as from wild type (Figure 4.6A lane 2 vs

lane 1), and the RNA that copurified was also comparable (Figure 4.6B lane 2 vs 1), indicating that while Clu1 mostly binds to $IP_6$ when it is present, in its absence, $IP_5$ is largely able to compensate. When Clu1 protein was purified from ipk2Δ yeast, very little protein was purified, and most of the protein purified was the MBP tag alone (Figure 4.6A lane 3) indicating that there is a problem with the expression or stability of the protein with only $IP_3$ is present. RNA binding was also greatly depleted in the protein purified from ipk2Δ yeast (Figure 4.6B lane 3), however this may be due to the instability of the protein, rather than a specific effect of $IP_6$ binding on RNA binding.

We also utilized size exclusion chromatography to analyze the Clu1 protein purified from the IP kinase null strains. This analysis confirmed the ability of $IP_5$ to compensate for $IP_6$ since the protein purified from WT and from ipk1Δ yeast were both eluted very early indicating that they were both stable and copurifying with RNA (Figure 4.6C solid line and large dashed line). This analysis also showed the the protein purified from ipk2Δ yeast was not bound to RNA, and interestingly that there was a population of protein that eluted at a volume consistant with a Clu1 monomer (Figure 4.6C small dashed line, and peak just before 13 ml), which has not been obsorved in any other condition. This indicates that $IP_6$ binding may be involved in the multimerization of Clu1, and that this multimerization may be required for both stability of the protein and RNA binding.

### 4.3.3 Size and Shape of the Clu1 Complex

Because the results from size exclusion chromatography of protein purified from ipk2Δ yeast showed $IP_6$ to be important for the multimerization state of Clu1, we sought to further characterize the composition of the Clu1 multimer. When Full length Clu1 is purified using a

size exclusion column it elutes near the void volume as a very high molecular weight complex of protein and RNA. When the RNA is removed either with RNAse or by deletion of the RNA binding domain, Clu1 elutes as a smaller, but still large protein complex, larger than its predicted molecular weight as a monomer (Figure 4.2A). For several $IP_6$ binding proteins $IP_6$ is involved in some way in mediating association of protein complexes ether as multiprotein complexes, or association of multiple subunits of the same protein. Since in the absence of more highly phosphorylated IPs Clu1 is destabilized and elutes from a size exclusion column at a volume consistent with a monomer, it seems likely that $IP_6$ may be involved in the association of multiple Clu1 monomers into a larger complex.

In order to better understand the way that $IP_6$ is regulating Clu1 we next sought to characterize the multimeric state of Clu1 by determining the size of the multimer. When Clu1 was purified as described above the size of the complex based on calibration of the Superdex 200 Increase 10/30 size exclusion column, was approximately 440kDa assuming a spherical globular protein. This result was also corroborated by dynamic light scattering analysis of Clu1ΔC50

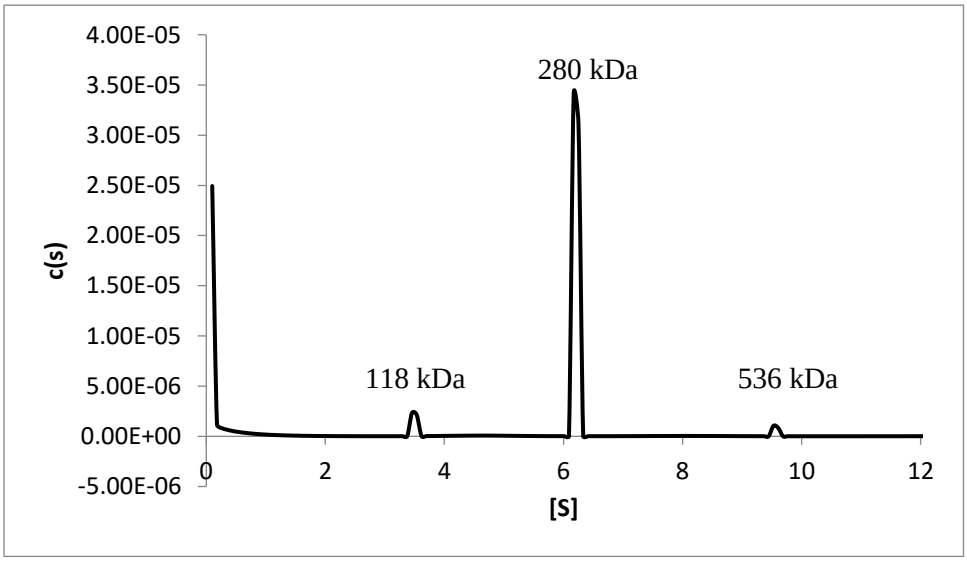


Figure 4.7 Analytical ultracentrifugation analysis of Clu1ΔC50. AUC analysis demonstrates that Clu1 is primarily a dimer, with small populations in a monomeric and tetrameric states.

protein which indicated a molecular weight of 436kDa for a spherical protein. Since the calculated molecular weight of a single Clu1ΔC50 monomer is 139kDa, these results would indicate that if Clu1 is a spherical protein it exists as a trimer; however, it is not known if Clu1 is spherical or slightly elongated. In order to better characterize the Clu1 multimer we utilized several techniques that would provide information about the shape of the complex as well. The first of these techniques utilized was analytical ultracentrifugation (AUC). This technique determined that Clu1 multimers predominantly have a molecular weight of 280kDa, consistent with a dimer, and small populations of protein with a molecular weight of 118kDa consistent with a monomer or a degradation fragment, and 536kDa consistent with a tetramer (Figure 4.7).

As a further characterization of Clu1 we also utilized single particle negative stain electron microscopy. The 2D class averages produced with this method also demonstrated an

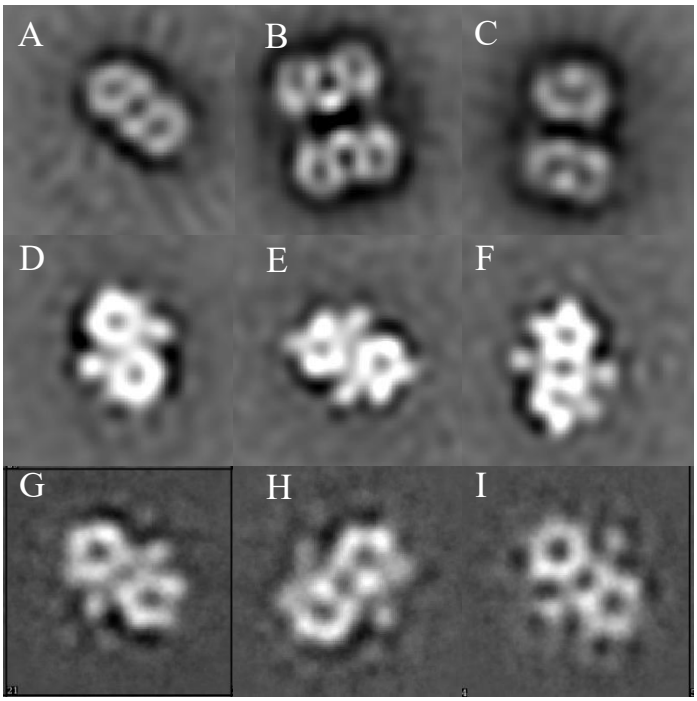


Figure 4.8 Representative 2D class averages of Clu1ΔC50. Negative Stain Class averages of Clu1Δ50. A-C) Class Averages from Clu1Δ50 initial data set. D-F) Class averages from Clu1Δ50 ipk2 null (with MBP tag). 2J-L Class averages from Clu1Δ50 with MBP tag

elongated particle, with symmetry that is consistent with a dimer as well (Figure 4.8A). Interestingly, there was also a class of protein that appeared to be an association of two dimers (Figure 4.8B), which may represent the tetramer population observed in AUC analysis. Utilizing negative stain electron microscopy, we examined Clu1ΔC50 protein purified from ipk2Δ to determine the role of $IP_6$ in the association of the Clu1 dimer subunits. When protein was purified from ipk2Δ yeast the protein was very unstable, and very little protein was able to be purified in the absence of the MBP tag. By leaving the MBP tag on the protein it seemed to be more stable, and this enabled the preparation of samples for electron microscopy with Clu1 from ipk2Δ yeast. As a control for any changes in the protein's structure caused by leaving the MBP tag attached Clu1Δ50 with MBP attached was also prepared from Wt yeast. The overall shape of the particles was the same as the Wt Clu1ΔC50 (Figure 4.8D-F), with the exception of two additional lobes around the center of the particle that were observed in both the wt purified and Δipk2 purified Clu1Δ50 with MBP tag (Figure 4.8G-I). These small lobes likely represent the 42kDa MBP fusion protein. Since they are oriented near the center of the protein, and the Clu1ΔC50 MBP fusion protein did not contain a linker between the tag and the N-terminus of Clu1, it is likely that the interface between the two subunits of the Clu1 dimer is very close to the N-terminus of the protein. In ipk2Δ purified Clu1 there were no classes of tetramer observed, however they were also not observed in the wt purified protein with the MBP tag left on either, so this may be an effect of the tag rather than of $IP_6$. Additionally, no classes that appeared to be a monomer were observed in any of the preparations. This seems to indicate that the Clu1 dimer can associate in the absence of $IP_6$, but that in the absence of $IP_6$ the dimer is less stable than when $IP_6$ is present.

Another technique used to analyze the multimeric state of Clu1 was native mass spectrometry. In this technique native protein complexes are transferred to the gas phase with electrospray ionization under conditions that preserve the structural conformation of the protein,

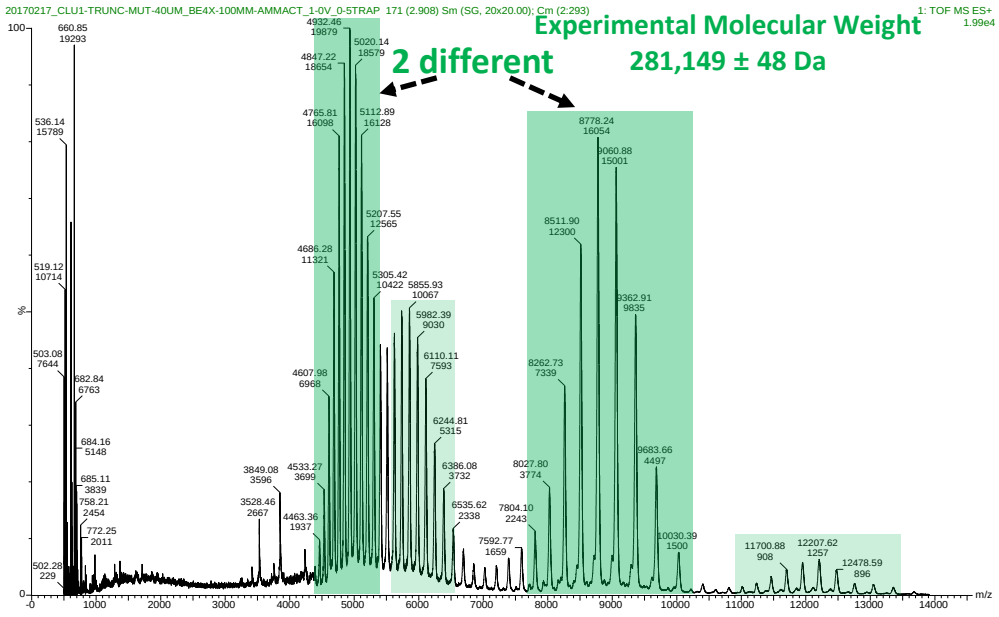


Figure 4.9 Native mass spectrometry analysis of Clu1ΔC50. Native Mass spectrometry confirms that Clu1 is a dimer, and that it exists in several different conformations highlighted in green.

and the now ionized protein complexes are then analyzed with mass spectrometry. This technique can provide information about the composition and stability of protein complexes, as well as different conformational changes within a complex. Analysis of Clu1ΔC50 showed that the Clu1 complex had a mass of 281kDa consistant with a dimer. The analysis also determined that there are at least two different major populations of protein (as well as at least two minor populations) with the same mass, but different mass to charge ratios (Figure 4.9). This indicates that Clu1 has at least two major different conformations that ionize differently when they enter the gas phase. Further energy-resolved conformational analysis of the different populations showed that while the two populations have different conformations, the different conformations have similar stability. These different popluations of Clu1 in different conformations will be

discussed furhter in the following chapter.

Collectively, this analysis of Clu1 shows, utilizing a variety of techniques, that Clu1 is a dimer with a slightly oblong shape. There is also some indication that these dimers may be able to associate into tetramers as well. Additionally, there are two conformations of Clu1 dimers that exhibit similar stability. It was also determined that while Clu1 dimers were able to form in the absence of $IP_6$, the resulting dimers were less stable than when $IP_6$ was present.

### 4.3.4 RNA Binding, $IP_6$ Binding and the Mitochondrial Clustering Phenotype

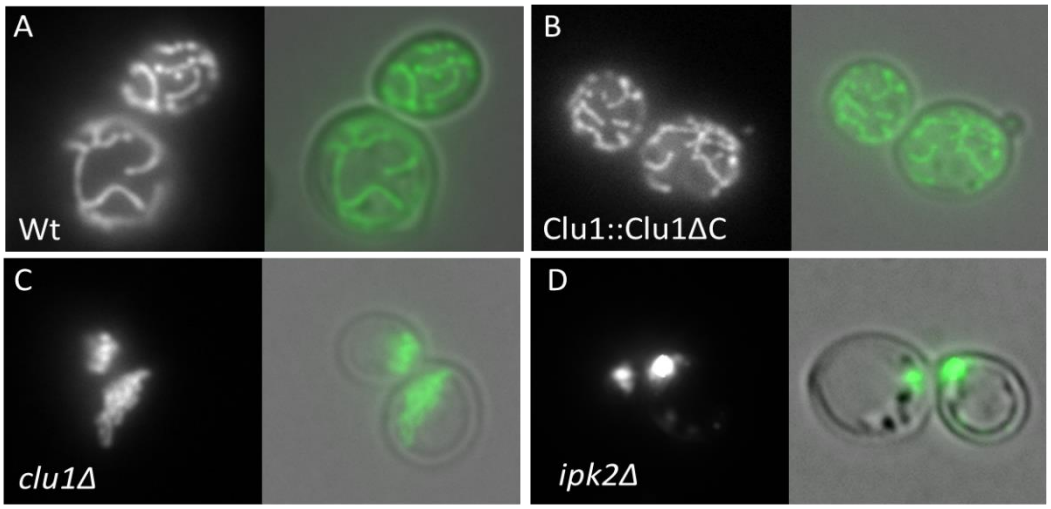


Figure 4.10 Microscopy analysis of the clustered mitochondria phenotype. Deletion of Clu1 or Ipk2 result in a clustered mitochondrial phenotype, but deletion of the C-terminus of Clu1 required for RNA binding exhibit distributed mitochondria. A) Wild type yeast exhibit mitochondria that are distributed throughout the cell, and B) yeast expressing Clu1 without the C-terminal RNA binding domain also have distributed mitochondria. C) clu1Δ yeast have mitochondria that are clustered together at the side of the cell, and D) yeast unable to produce $IP_6$ have a severe clustered mitochondrial phenotype.

The primary phenotype associated with loss of Clu1 protein across all the species where it has been observed is a clustering of the mitochondria (Figure 4.10C vs A). The binding of $IP_6$ to Clu1 has previously been shown to be required for correct mitochondrial localization (Pham, 2012), and ipk2Δ yeast have an even more severe mitochondrial clustering phenotype than clu1Δ yeast (Figure10D) (Dimmer *et al.*, 2002). In order to determine the role that RNA binding plays

in this mitochondrial phenotype microscopy experiments were conducted on Wt yeast as well as yeast with truncated Clu1 protein (Clu1::Clu1ΔC). It was observed that while clu1Δ yeast display the typical clustered phenotype (Figure 4.10C), Clu1::Clu1ΔC yeast expressing the truncated form of the protein, which has been shown to not bind RNA, display a wildtype mitochondrial phenotype (Figure 4.10B). This demonstrates that while $IP_6$ is required for proper mitochondrial disrtrbution, RNA binding is not required. Further investigation of Clu1 may need to focus on determing which functions of Clu1 are dependent on RNA binding, and which ones are separate. Since Clu1 is a multidomain protein, it is possible that different domains may have different functions.

### 4.3.5 Other Clu1 Phenotypes

In addition to the clustered mitochondria phenotype, Clu1 homologues have also been implicated to have other phenotypes in other systems, notably the post birth lethality observed in Cluh knockout mice that was associated with the inability to switch metabolic states (Schatton and Rugarli, 2018). In order to determine if this role in the regulation of metabolism in response to altered nutritional states, we sought to characterize the ability of clu1Δ yeast to respond to nutritional changes. Normally in a laboratory setting yeast are grown at 30°C on rich media with 2% Dextrose as the primary carbon source. We observed that when grown at 37°C on media with much less dextrose, 0.1% instead of the normal 2%, clu1Δ yeast had slightly less growth, and ipk2Δ yeast were unable to grow (Figure 4.11A). In order to further investigate the role of Clu1 in regulation of utilization of alternate carbon sources we grew yeast on media containing the non-fermentable carbon sources ethanol, lactate and glycerol. On media containing Ethanol, but not the other non-fermentable carbon sources clu1Δ yeast exhibited impaired growth (Figure

4.11B). Since growth was only inhibited on Ethanol, and not on the other carbon sources, we decided to determine if this was due to inhibition by ethanol rather than by having to utilize an

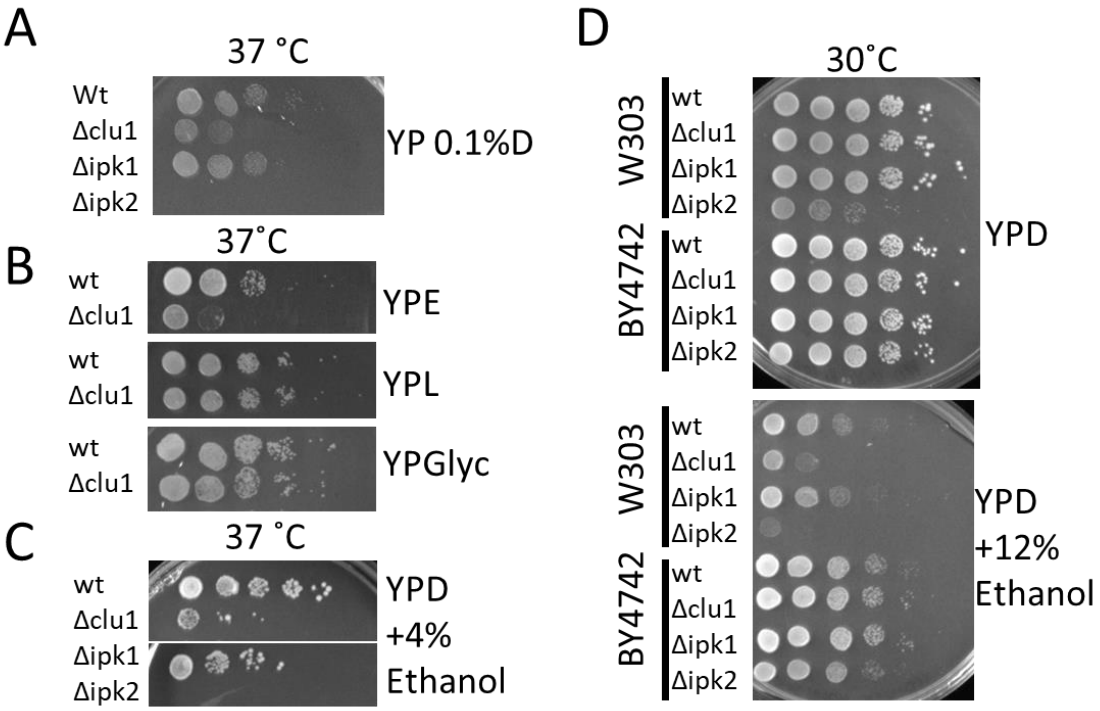


Figure 4.11 Clu1 is required for growth on ethanol. Clu1 and IP kinase mutants have growth defects on altered carbon sources. A) When grown on media with minimal dextrose Δclu1 yeast and Δipk2 yeast show decreased growth. B) On media with the non-fermentable carbon sources Glycerol and Lactate Δclu1 yeast grow similarly to wild type, but on media with ethanol as the sole carbon source Δclu1 yeast have impaired growth. C) Δclu1 and Δipk2 yeast have increased ethanol sensitivity. D) The ethanol sensitivity of Δclu1 and Δipk2 yeast is specific to the W303 background, as BYY4742 Δclu1 and Δipk2 yeast grow similarly to wild type.

alternative carbon source, and indeed clu1Δ yeast were less competitive when grown on media containing both ethanol and dextrose, and ipk2Δ yeast were unable to grow at all (Figure 4.11C). Since yeast are primary commercial producers of ethanol, there is much study of ethanol tolerance in yeast (Snoek, Verstrepen and Voordeckers, 2016), and none of these studies had identified Clu1 as a gene involved in ethanol tolerance. One reason for this may be due to differences between strains utilized in screens often yeast strains from breweries and wineries

compared to BY4742, and the W303 strain characterized here. To test this clu1Δ and IP kinase null yeast from both w303 background as well as from BY4742, a common strain used in high throughput screens, were grown on ethanol. This result confirmed that clu1Δ and ipk2Δ yeast from w303 are both sensitive to ethanol, but that in the BY4742 background they showed no increased sensitivity to ethanol (Figure 4.11D).

The *Drosophila* Clu1 homologue Clueless has been identified as being involved in clearing damaged mitochondria after oxidative damage (Wang, Clark and Geisbrecht, 2016). In order to test this in yeast, clu1Δ and IP kinase null yeast were grown on media containing the mitochondrial depolarization agent Carbonyl cyanide m-chlorophenyl hydrazone (CCCP). While all the strains grew similarly at 30°C, clu1Δ and ipk1Δ yeast had decreased growth in the presence of CCCP at 37C, and ipk2Δ yeast were completely unable to grow (Figure 4.12). This indicates that the role for Clu1 in responding to mitochondrial damage is conserved in yeast as well.

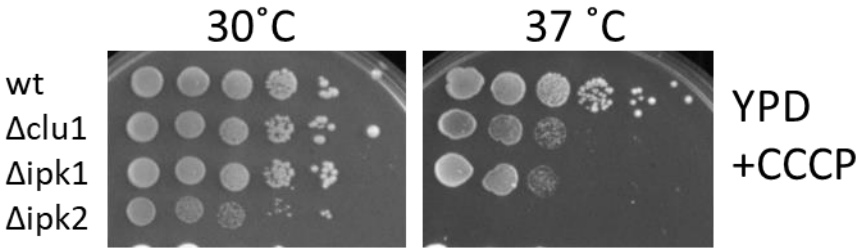


Figure 4.12 Clu1 is required for growth with mitochondrial depolarization. Clu1 null yeast are sensitive to mitochondrial depolarization agent CCCP at 37°C. At 30°C Δclu1 and Δipk2 grow like wild type in the presence of CCCP, but at 37°C Δclu1, Δipk1 and Δipk2 strains all shown decreased growth compared to wild type, with Δipk2 unable to grow at all.

## 4.4 Discussion

Clu1 is a protein that is primarily characterized for its role in the proper distribution of

mitochondria. It has also been shown to play roles both in the response to damaged mitochondria, as well as in adapting the metabolic state of a cell in response to metabolic changes. One of the proposed mechanisms for this is through its binding to mRNAs that encode mitochondrial proteins. It has also been identified as a protein that binds to $IP_6$ as a structural cofactor. In this chapter we show that the C-terminus of the protein is required for binding to RNA, and that specifically the C-terminal 50 amino acids are necessary and sufficient for mediating this interaction. We also show that its RNA binding is independent from its $IP_6$ binding. While $IP_6$ is the primary IP bound by Clu1, we show that $IP_5$ can also bind, and that in the absence of $IP_6$, $IP_5$ can largely compensate. We also show that while $IP_5$ can rescue the function of $IP_6$, $IP_3$ cannot. Clu1 purified from ipk2Δ yeast exhibit less stability, and less RNA binding. We also show that Clu1 is bound to $IP_6$ at a ratio of 1:1, and Clu1 exists primarily as a dimer that is slightly elongated in shape. The Clu1 dimer exists in two conformations that exhibit similar stability, and a Clu1 tetramer can form under certain circumstances. With size exclusion analysis of Clu1 protein purified from ipk2Δ yeast we observe both dimer and monomer Clu1, however other forms of analysis of Clu1 protein from ipk2Δ yeast have only shown a dimer. This analysis is quite difficult due to the instability of the protein purified under these conditions. We also characterize the phenotype of clu1Δ yeast. While Clu1s' RNA binding is not required for proper mitochondrial localization, the presence of $IP_6$ is required. We also confirm that the requirement of Clu1 for response to changing metabolic states, and response to mitochondrial damage observed in other organisms are conserved to yeast as well. Here we have provided a characterization of Clu1 both on a protein level, and phenotypically. The next chapter will discuss the structural characterization of Clu1.

# Chapter 5 Structural Characterization of Clu1

## 5.1 Introduction

Clu1 is a protein that has been identified as playing a role in mitochondrial distribution, RNA binding, as well as being an $IP_6$ binding protein. While several $IP_6$ binding proteins were discovered when their crystal structures were solved, and $IP_6$ was identified within their structure, Clu1 was identified through a targeted screen for $IP_6$ binding proteins, and other than the data presented in Chapter 4, very little is known about its structure. One structural element that it shares a few other $IP_6$ binding proteins is the presence of a tandem repeat domain (Clu1 domain architecture shown in Figure 5.1). $IP_6$ binding mediated by tandem repeat domains has been identified in Coi1 (Sheard *et al.*, 2010) and Tir1 (Tan *et al.*, 2007), with their Leucine Rich Repeat (LRR) domains, in Pds5 (Ouyang *et al.*, 2016a) with its HEAT repeat (named for the proteins where it was initially described Huntingtin, elongation factor 3 (EF3), protein phosphatase 2A (PP2A), and TOR1), and in NatA and NatE with their tetratricopeptide repeat (TPR) domains (Neubauer, 2012; Pham, 2012). In each of these cases $IP_6$ binds to the center of the super solenoid ring formed by the tandem repeat domains.

TPR domains are formed by a loosely conserved sequence of 34 amino acids that follow

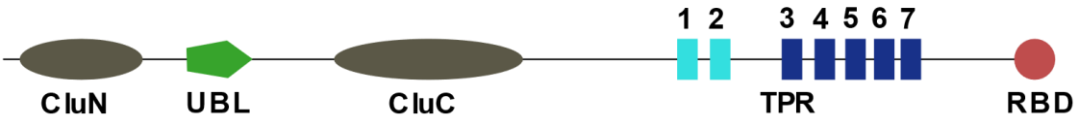


Figure 5.1. Domain organization of Clu1. Clu1 contains Clu1 N-terminal and Clu1 Central domains that are conserved throughout the Clu1 protein family but are not homologous to any other protein domains. Between the Clu-N and Clu-Central domains Clu1 contains a Ubiquitin Like Domain (UBL). Clu1 $IP_6$ binding is mediated by its TPR domains, 5 of which are predicted by SMART and TPRPRED, and the two $IP_6$ binding TPRs (identified by Pham 2012), that were manually annotated. Clu1 also contains an RNA binding domain that was identified and described in Chapter 4.

the pattern of consensus at eight different positions throughout the domain: (Trp/Leu/Phe) 4, (Leu/Ile/Met) 7, (Gly/Ala/Ser) 8, (Tyr/Leu/Phe) 11, (Ala/Ser/Glu) 20, (Phe/Tyr/Leu) 24, (Ala/Ser/Leu) 27, and (Pro/Lys/Glu) at 32. This pattern however is degenerate, and therefore difficult to predict. What is conserved throughout TPRs is the structure. TPRs consist of arrays of two helix bundles, where each pair loosely contains the conserved sequence. One of the TPR repeat two helix bundles from NatA is shown in Figure 5.2A. The arrays of helix pairs are assembled into a larger super solenoid. The TPR supersolenoid composed of 13 TPR repeats from NatA is shown in Figure 5.2B. Often, TPR domains are involved in mediating protein-protein interactions. Since they can be difficult to predict based on sequence alone, the family of TPR containing proteins is continuing to grow (D'Andrea and Regan, 2003).

Clu1 contains several TPR repeats like NatA, and 5 of these domains are predicted by SMART (Letunic and Bork, 2018), and an additional two domains can be manually annotated (Pham, 2012). This is not surprising since TPR domains are not highly sequence conserved,

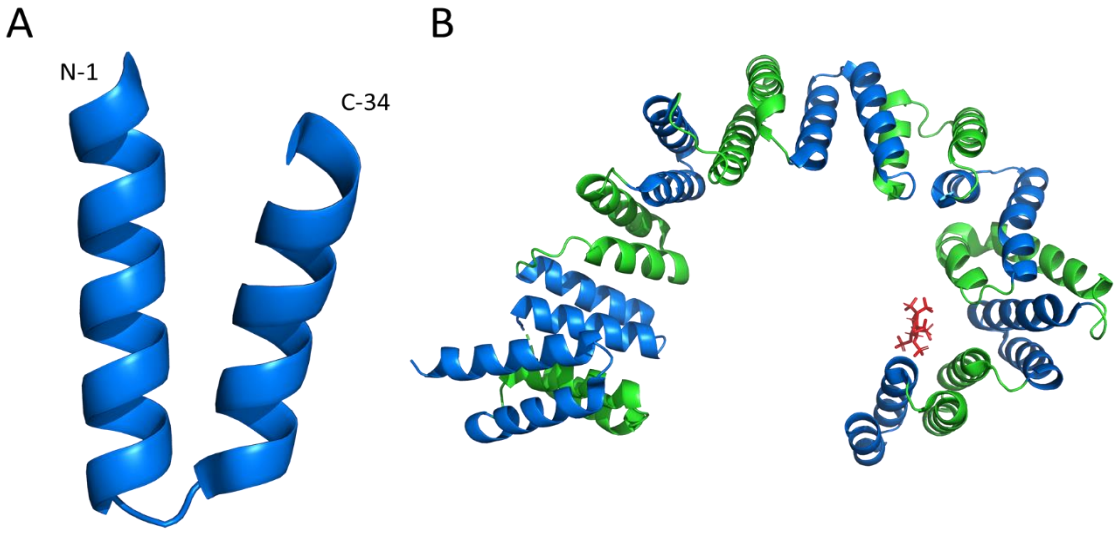


Figure 5.2 Representation of TPR repeat, and TPR solenoid domain. TPR shown bound to $IP_6$ from NatA (PDB ID 4HNW). A) A single TPR repeat consisting of two α helices composed of 34 amino acids from NatA TPR 1. B) The TPR domain super solenoid composed of 13 TPR repeats from NatA, with $IP_6$ bound to TPR12 and TPR13.

rather it is a structural conservation. This was also observed with the structure of NatA was solved, as it contained 7 predicted TPR repeats, but the solved structure contained 15 TPR repeats (Neubauer, 2012). The $IP_6$ binding site in Clu1 and NatA were both predicted based on a repeating motif of positively charged amino acids facing the same direction on the helix bundles of the repeats and this approach was validated by the structure of NatA. Based on mutation of the predicted residues in Clu1, the $IP_6$ binding site was determined to be mediated by the two manually annotated TPR repeats (Pham, 2012).

In addition to TPR repeats, Clu1 is also predicted to contain Clu domains, the CluN terminal domain and the CluCentral domain (Figure 5.1). These domains are highly conserved in Clu1 homologues across eukaryotic species but are not found in any proteins outside of the Clu1 family based on analysis with PFAM (El-Gebali *et al.*, 2019). As none of the Clu1 family proteins have been characterized structurally, very little is known about these domains. Between the CluN and CluCentral domains is a portion of the protein that based on secondary structure prediction has similarity to a Ubiquitin like domain (Pham, 2012), however the functional significance of this domain is also not known. Finally, the RNA binding of Clu1 was identified to take place via a series of positively charged amino acids at the C-terminus of the protein as described in Chapter 4, defining the C-terminus as an RNA binding domain. In order to better understand how Clu1 is functioning in cells to bind to RNA as well as other proteins, this chapter describes further structural characterization of the Clu1 protein.

## 5.2 Methods

### 5.2.1 Growth of Yeast and Protein Purification

Growth of yeast and purification of Clu1ΔC50 protein was conducted as described in Chapter 4.

### 5.2.1 Negative Stain Electron Microscopy

To prepare grids, 4 ul of 0.01 mg/ml Clu1, or 0.05 mg/ml of refolded Clu1 was applied to grids, for 60 seconds. Excess protein was removed by blotting with filter paper. The grids were then washed 2x with water, and then once with Uranyl formate solution (0.7% w/v), and then stained with Uranyl formate solution for 90 seconds. Excess stain was removed by blotting and aspiration. Negative stain data was collected at 200kV with a FEI Tecani TF20 with a CCD camera. Images were collected at 62000x and a pixel size of 1.757Å. Negative stain data was analyzed with the EMAN2 software package (Tang *et al.*, 2007). Semi-automated particle picking, and 2d class averages, as well as construction of a low-resolution initial model were all conducted in EMAN2. Processing of data from refolded protein was conducted with Scipion (de la Rosa-Trevín *et al.*, 2016) utilizing programs from the Xmipp (Sorzano *et al.*, 2004) and Relion (Scheres, 2012) software packages.

### 5.2.3 Circular Dichroism Analysis of Clu1 and Refolding Assays

Circular dichroism (CD) spectra of Clu1ΔC50 was measured with a Jasco J-840 spectropolarimeter. Spectra between 200-250 nm were measured in 1 mm pathlength cell with 5-times scanning and scanning speed 10 nm/min with 1 nm interval data collection. Protein concentration was 0.5 mg/ml. Collected spectra were analyzed with CAPITO program

(Wiedemann, Bellstedt and Görlach, 2013). For unfolding of Clu1 with Urea, 0.5 mg/ml of Clu1 was denatured with 8 M urea in 20 mM Tris pH 7.5 50 mM NaCl for 30 min at room temperature. The denatured protein and urea were then diluted in steps from 8 M, to 6 M, 5 M, 4 M, 3 M, 2 M, 1 M, 0.5 M steps with 30 min of incubation for each of the first 4 steps, and then 1 hour incubation for each of the final 4 steps. Removal of $IP_6$ from unfolded protein was done using $TiO_2$ beads based on the method described by (Wilson *et al.*, 2015). The removal of $IP_6$ from the sample was verified with a Vip kinase reaction as described previously. Samples were diluted using 50 mM Tris pH 7.5 and 150 mM NaCl, either with or without a molar excess of $IP_6$ added as indicated. The refolded protein was concentrated with 10k MWCO spin concentration devices from Millipore before analysis with CD or negative stain electron microscopy.

### 5.2.4 Cryo Electron Microscopy Analysis of Clu1

For single particle electron microscopy analysis of Clu1 grids were prepared using a FEI vitrobot mark IV. Quantifoil ultra AU gold foil grids with 300 mesh and 1.2/1.3 spacing were glow discharged for 20 seconds. Then, 4 ul of sample was applied for 1 seconds at 21°C and 75% humidity, then blotted for 4.5 seconds with a blot force of 1, then after a 0.5 second delay the grids were flash frozen in liquid ethane, and transferred to grid boxes in liquid nitrogen and stored in liquid nitrogen until used. Data collection was conducted with a Polara TF30 operating at 300kV. Data was preprocessed using MotionCorr (Li *et al.*, 2013) and gCTF (Zhang, 2016). Further data analysis was conducted using Relion (Scheres, 2012). Initially, a small set of about 1000 particles was picked manually from a variety of different quality micrographs from each data set collected. These particles were 2D class averaged, and the resulting 2D Averages were used for automated picking for the remainder of the dataset.

### 5.2.5 X-Ray Crystallography

For crystallization protein purified as described previously was concentrated, usually to about 15 mg/ml and the buffer was changed to 20 mM Tris, pH 8.0, 20 mM NaCl 1 mM TCEP. It was possible to concentrate protein to about 50 mg/ml and use that concentrated sample for crystallization with good crystal growth. The sitting drop method of crystallization was used with Rigaku CombiClover 4 Chamber 4x250 Crystallization Plates. For crystallization setup protein solution was mixed with precipitant with protein:precipitant ratios of 1:1, 2:1 and 1:2 v/v. The protein concentration was varied in the crystallization trials as well as different temperature conditions with room temperature as the starting conditions as well as 20°C and 12°C.

## 5.3 Results

### 5.3.1 Circular Dichroism and Clu1 Refolding

In order to establish the secondary structure characteristics of Clu1 we sought to utilize Circular Dichroism (CD) to analyze Clu1. CD is a method of analysis based on the principle that proteins and other optically active molecules differentially absorb circularly polarized light based on their structural conformation. Using this technique one can measure the secondary structure that a particular protein has based on how it absorbs the circularly polarized light, with the output unit of ellipticity being correct for the concentration of the protein (resulting in molar ellipticity as the final unit). Circular Dichroism analysis established that Clu1 is largely α helical. Based on analysis of the molar ellipticity with the program CAPITO (Wiedemann, Bellstedt and

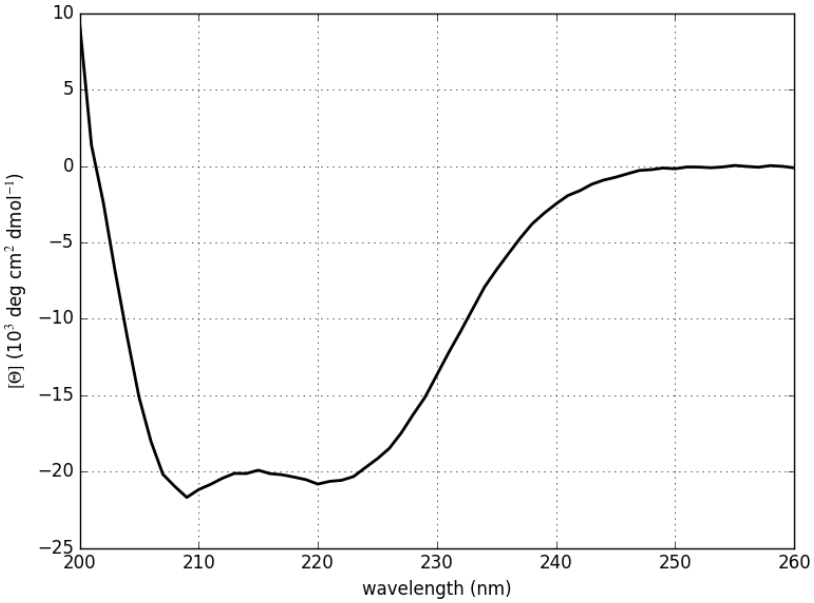


Figure 5.3 Circular Dichroism analysis of Clu1. This analysis indicates Clu1 contains fairly typical secondary structure for a largely globular protein and consists mostly of α helix.

Görlach, 2013) the secondary structure of Clu1 is 57% alpha helical, 15% beta sheet and the remainder is unstructured.

Next, we wanted to utilize protein unfolding with urea and refolding by diluting out the urea, both with and without $IP_6$ as a method to characterize the role that $IP_6$ plays in the structure of Clu1. If $IP_6$ is a factor required for Clu1 folding, after denaturation, Clu1 should only be able to refold in the presence of $IP_6$. In order to test this, a method for removing the endogenous $IP_6$ folded inside of Clu1 needed to be established, and to accomplish this $TiO_2$ beads were utilized. Clu1ΔC50 protein that had been denatured in 8M Urea was incubated with 4 mg of $TiO_2$ beads. After the beads were removed, a Vip kinase assay was utilized to confirm that $IP_6$ had been removed from the samples that had been incubated with beads (Figure 5.4A). As a control for $IP_6$ depletion, $IP_6$ in buffer was also incubated with $TiO_2$ beads (Figure 5.4A). For refolding the following conditions were compared, native Clu1ΔC50 without any denaturation, Clu1ΔC50 denatured in 8M urea and not refolded, Clu1ΔC50 denatured and no $IP_6$ removed (this will

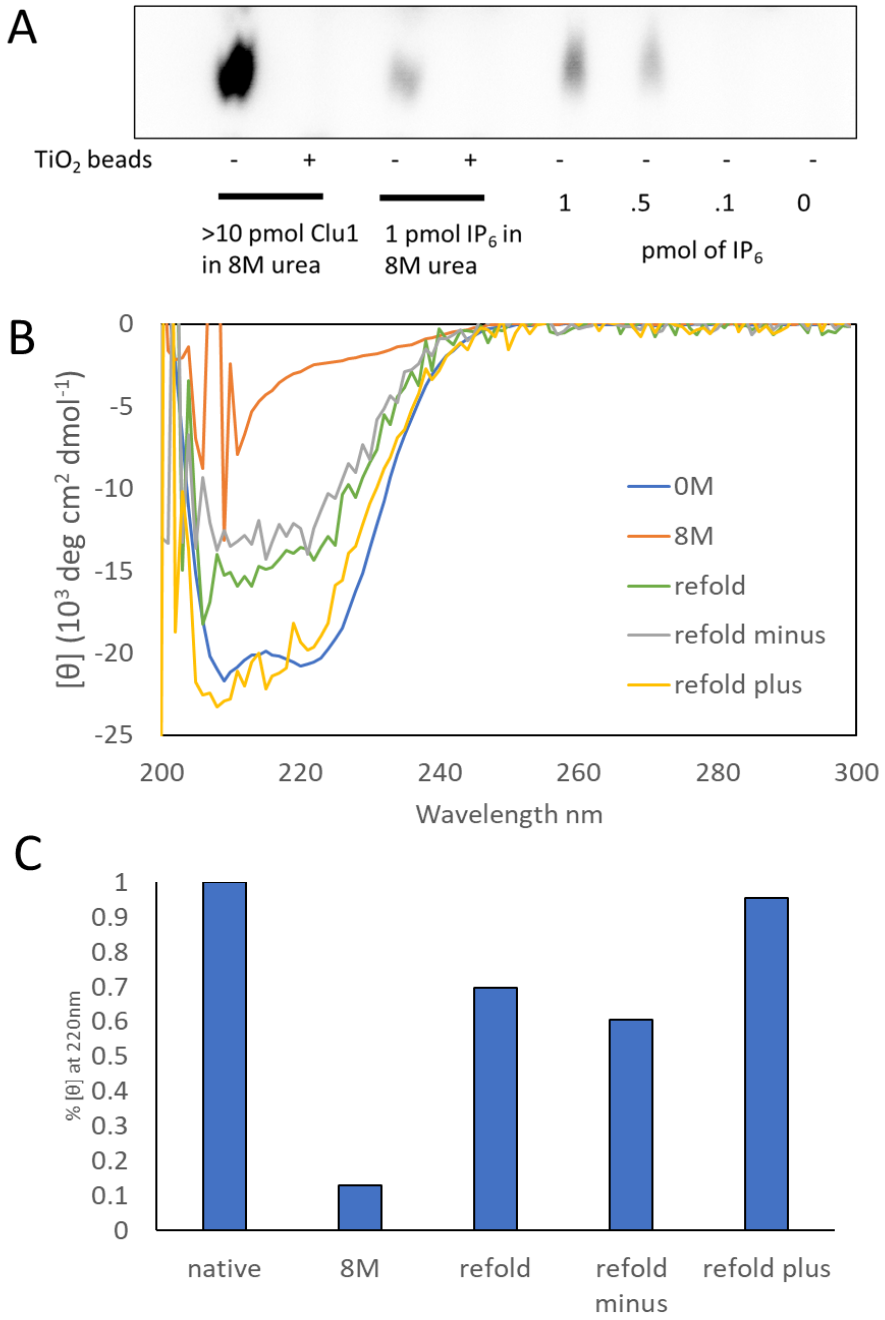


Figure 5.4 Denaturation and refolding of Clu1. Urea denaturation and subsequent $IP_6$ dependent refolding of Clu1. A) $IP_6$ was removed from denatured Clu1 by incubation with $TiO_2$ beads and confirmed with Vip kinase assays. B-C) Extent of refolding was determined by analysis with CD.

contain 1 Molar equivalent of $IP_6$), Clu1ΔC50 denatured in 8M urea with $IP_6$ removed but no $IP_6$ added back, and Clu1ΔC50 with $IP_6$ removed and 10x molar excess of $IP_6$ added back during refolding. When no $IP_6$ is added back Clu1 is unable to completely refold as indicated by CD analysis showing it achieves only about 60% minimum molar ellipticity at 220 nm (Figure 5.4B and C refold minus). When $IP_6$ is not removed from the unfolded protein, leaving behind 1 molar equivalent of $IP_6$ Clu1 is able to refold marginally better, but still only achieving about 70% of minimum molar ellipticity at 220 nm (Figure 5.4B and C refold). When 10x molar excess of $IP_6$ is added to Clu1 during refolding it is able to refold nearly completely, reaching about 95% minimum molar ellipticity at 220 nm (Figure 5.4B and C refold plus). This demonstrates that $IP_6$ is required for reestablishment of secondary structure during Clu1 refolding. The overall CD profile of the refolded plus $IP_6$ protein was similar to the native protein, but all of the denatured samples had much noisier CD profiles when compared to native protein. This was likely due to loss of sample during the unfolding and dilution process. Yield of refolded protein was increased with increasing amounts of $IP_6$ added back during refolding, further demonstrating the role of $IP_6$ in the folding of the protein.

In order to confirm that the Clu1 in the refolded sample was refolded into a native state, and not merely a partial restoration of certain secondary structural elements further analysis was needed. To do this Clu1 that was refolded in the presence of 10x molar excess $IP_6$ was analyzed with single particle negative stain electron microscopy. This analysis revealed that the refolded Clu1 was very similar to native Clu1 protein with the overall size and shape of the particles being the same (Figure 5.5 lower compared to upper). One difference between the refolded protein and the native protein was that in the refolded protein many of the classes were less symmetrical than in the native protein. While some of the classes exhibited similar degrees of

symmetry to the native protein, many of the classes appeared to have an extended portion on one side. Also, in the refolded protein no tetramers were observed.

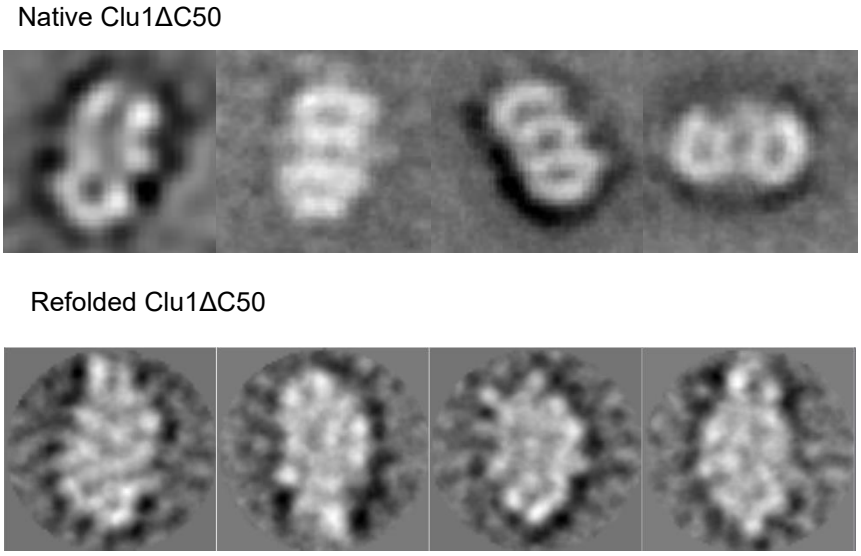


Figure 5.5 2D class averages from refolded Clu1. When Clu1ΔC50 is denatured and then refolded in the presence of $IP_6$ (lower) has a similar overall structure to 2D class averages from native Clu1.

### 5.3.2 Cryo Electron Microscopy

To further characterize the structure of Clu1, single particle cryo electron microscopy was utilized. In order to utilize this technique, conditions needed to be determined for preparation of Clu1 in vitrified ice. To do this a wide variety of conditions were tested primarily using Quantifoil copper grids with a variety of mesh sizes, and with a variety of protein concentrations, and blotting conditions. In all the conditions tested no Clu1 particles were observed in the vitrified ice using the carbon coated copper grids. It was however observed that there appeared to be particles on the carbon support around the edges of the ice. Based on this we hypothesized that Clu1 preferred to interact with the carbon support rather than with the vitrified ice. In order to test this, we attempted to determine conditions for grid preparation with

Quantifoil ultra Au gold foil grids. These grids are made of a gold mesh support instead of copper support, and rather than a carbon film with holes for vitrified ice they have a gold foil with holes for the vitrified ice. Immediately it was clear that these gold grids without carbon were much better for preparation of Clu1, and many particles were observed in the ice. Blotting conditions were determined, and grids with about 250 particles per image were prepared.

With samples prepared on these gold grids, an initial data set was collected and processed. This initial data set contained about 60,000 particles that were selected using a combination of manual and semiautomated particle picking. The appearance of the particles was very similar to the single elongated particles from the negative stain experiments, and only dimers were observed with no tetramers or monomers present. After cleaning up the automated particle picking, 2D class averaging was conducted narrowing the set to about 45,000. These initial 2D classes had a resolution between 10Å and 6Å (based on Relion estimates) and they included two main conformations (Figure 5.6A and B). One conformation has two closed ends (Figure 5.6B) and was largely symmetrical, and the other one has a more extended section on one of the ends (Figure 5.6A. These two classes may represent the different conformations that

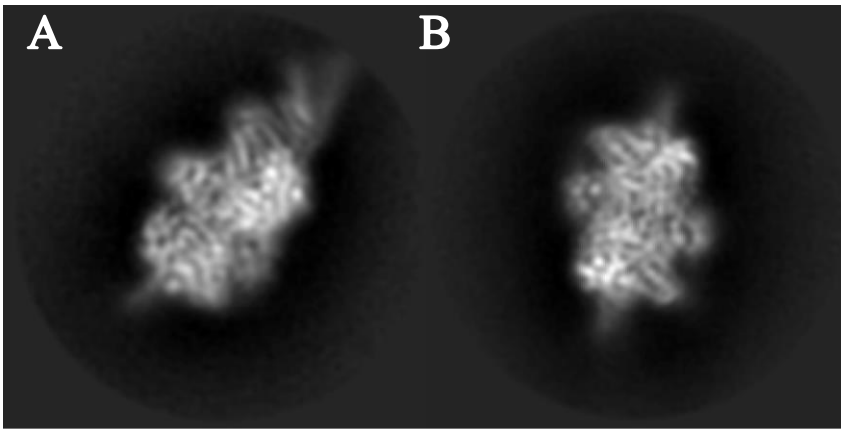


Figure 5.6. 2D class averages from Cryo EM analysis of Clu1. Both A and B represent 2D averages with roughly 6Å resolution. The class represented in A shows a more open end than the class represented in B.

were detected by native mass spectrometry, and while the differences are not discernable in the particles during picking, they are easily separable during 2D class averaging. The distribution of the particles between the two classes generally seemed to be about even. Interestingly no class was observed with the open extended portion on both ends as might be expected. The reason for this is unknown, but the possibility that the lack on an "open-open" Clu1 conformation, may be important to its biological function, or to the regulation of the protein.

After 2D classification the top 45,000 particles were selected and 3D classification was conducted. As a basis for classification an initial 3D model was required. This initial model was generated from the negative stain particles corresponding to the dimer class and was down sampled to 60Å. The 3D classification was conducted with 8 classes, but 80% of the particles were sorted into 2 classes, and the distribution of the particles was about even between the two classes with the "open" class (Figure 5.7A) containing about 18,700 particles and the "closed" class (Figure 5.7B) containing about 18,600 particles.

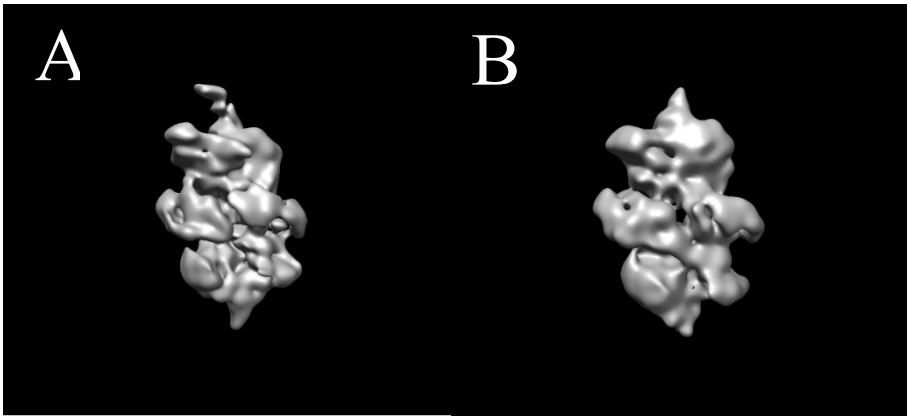


Figure 5.7. 3D classes from Cryo EM analysis of Clu1. A) represents the "open" class and B) the "closed" class

Since the classes were split fairly evenly further processing of the classes separately was not possible, since the data set was effectively cut in half. In order to further refine the data, the classes were masked, removing the extended portion of the protein where the majority of difference between the classes, and the particles were pooled, and refined. While the resulting 3D model (Figure 5.8A) is relatively low resolution (Relion estimate 8.2Å), it provides some useful information about Clu1's structure. One of the striking features is the curved ridge that

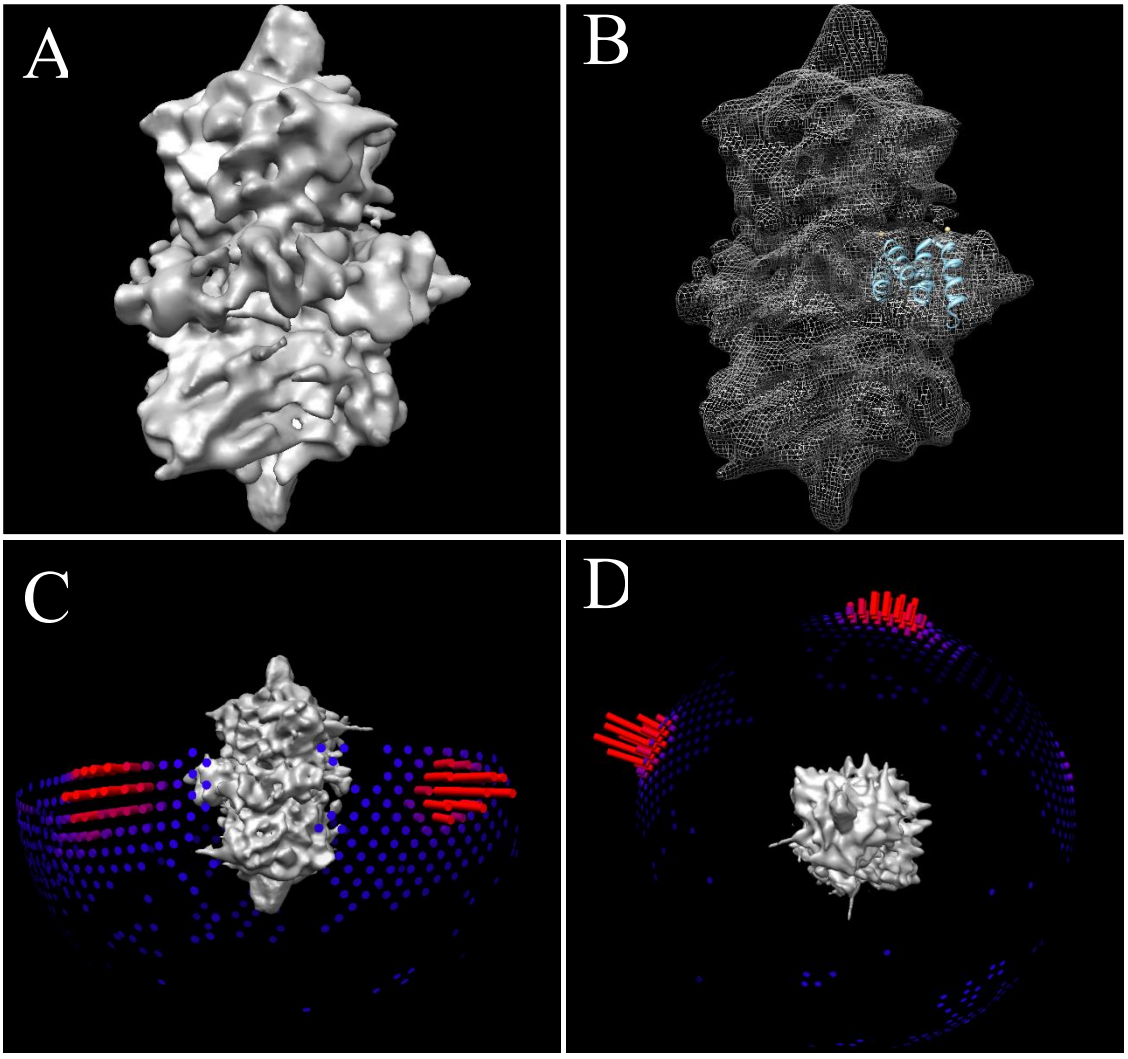


Figure 5.8. 3D Model of Clu1 from initial CryoEM data collection. Data is from pooling of both major 3D classes. A) surface model and B) mesh with 2 TPR repeats docked in (TPR structure is from X-Ray structure of consensus TPR, (PDB ID 2AVP), Kajander et. al, 2005) C) and D) show the Euler sphere of represented views from the site and top views

stretches around the center of the particle. Based on docking TPR domains from the crystal structure of the consensus TPR sequence (Kajander et. al, 2005), the dimensions of this ring are consistent with a solenoid of TPR repeats curving around the structure, with the remaining repeats sitting at either end of the protein (Figure 5.8B). If this band around the center is the TPR solenoid, then $IP_6$ would be bound near the interface between the two subunits of the dimer and may play an important role in the mediating the interactions between the subunits of the Clu1 dimer. It is also apparent from this model, that there are some serious distortions. By examining the orientations of the particles that make up the model, it is apparent that the distortions are caused by the limited number of orientations present in the dataset (Figure 5.8C and D). One possible reason for this is simply the limited size of the dataset. To address this, additional data was collected. Generally, the resolution improved with additional data (Relion estimate 6.4Å), but the distortions and limited orientations were still apparent.

To address this preferred orientation problem, one commonly utilized approach is to introduce mild detergents into the buffer prior to application of the protein to grids and vitrification of the sample in order to change the interactions of the protein with the air water interface and introduce new orientations of the protein. The changes in the buffer can cause problems with sample interaction with the grid, proper vitrification of the sample, as well as problems with resolution. A similar approach is to utilize Amphipol an amphipathic polymer that can provide similar changes to the interaction of the protein with the buffer. After screening through a variety of conditions with various detergents and Amphipol, a suitable condition was unable to be determined. All of the detergent conditions tried resulted in ice without any visible particles.

Another approach to address preferred orientations of a protein is to collect data with the

microscope stage tilted relative to the electron beam to expose new faces of the protein while it is still in the same preferred orientation relative to the plane of the ice. This approach can also introduce problems both with data collection in the form of the stability of the beam, as well as with data processing as the particles from different positions across each image will have different defocus values. To address beam induced sample movement gold grids can be utilized as they have been observed to minimize interference of the beam with the grid support. To address defocus issues, CTF correction can be conducted on a per particle basis rather than on a per image basis and thus correct the issue introduced by the tilted stage.

In order to determine if tilting could be of benefit in this case, and if so what tilt would be required to address problems with distorting of the map a program called CryoEF was utilized (Naydenova and Russo, 2017). CryoEF is a program that was developed to assess CryoEM data sets and determine the degree of distortion that was introduced into a data set based on the particles orientations present in the model. The key readouts from CryoEF include a recommended collection tilt angle, and an efficiency number. Efficiency numbers reflect how well covered the model is with the particle orientations present. If the efficiency number for a given data set is below 0.8, the program predicts that it is impossible to solve the structure due to distortions from the limited orientations. When the data collected for Clu1 without any stage tilt was examined with CryoEF, the efficiency number was 0.41, and the recommended angle of collection to address the preferred orientation bias was between 34° and 42°.

Two data sets were collected utilizing a tilted stage, the first tilt was limited to 15° due to issues with microscope stability, and the second dataset was collected at 35°. The data was then combined with the previously collected data. The combined data set is composed of 3,812 micrographs, and 753,122 particles. All micrographs were Motion corrected using MotionCorr

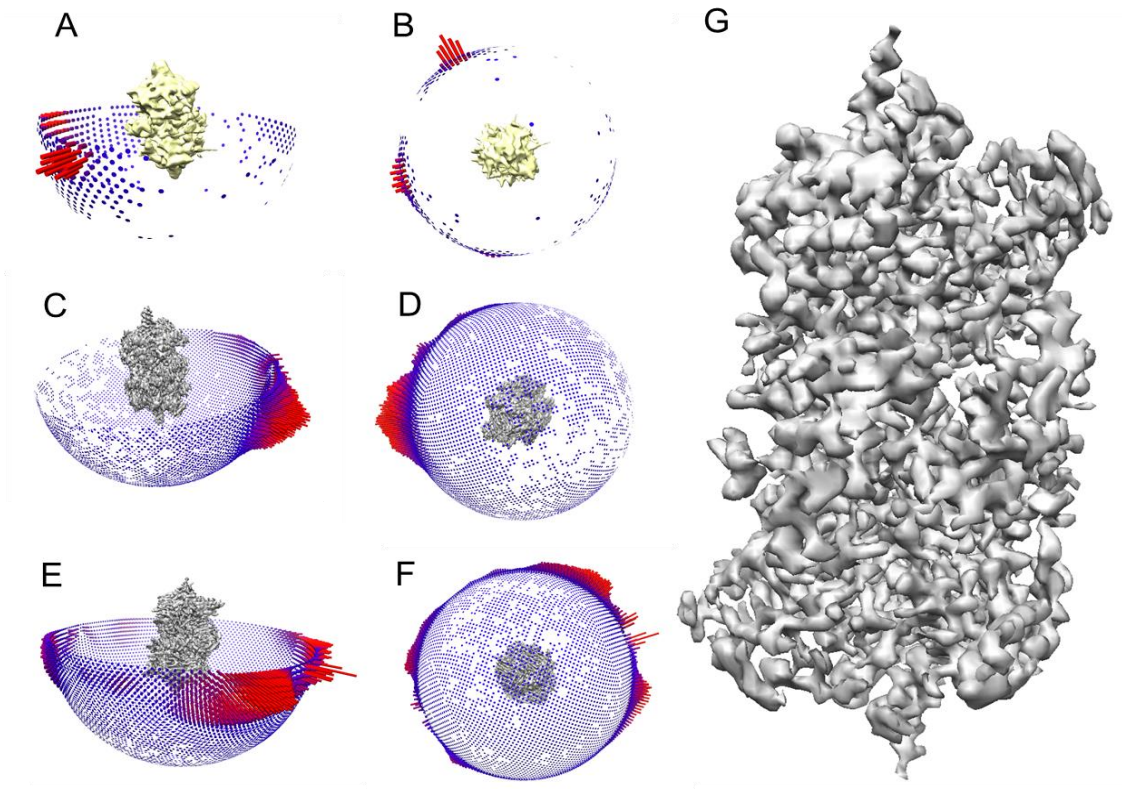


Figure 5.9 Collection of tilt data improves but does not fix the problem with preferred orientation. Orientations of the initial model (A) and (B). Orientations of tilt collected data (C) and (D). Orientations of Tilt data pooled with previous data (E) and (F). 4.2Å Map of Clu1 dimer (G).

2.0 and were then CTF corrected with gCTF. For images collected with a tilted stage the picked particles were corrected again using gCTF on a per particle basis and recombined with particles collected without tilt. Only images with an estimated resolution of 5Å or better were kept for further processing. All further processing was done using Relion 2.1 and particles were picked as before. These particles were subjected to 2 rounds of 2D class averaging, selecting only classes with visible secondary structure. These particles were then 3D classified once without symmetry to separate the two major conformations ("open" and "closed"), and each of these was processed separately. The "open" class and "closed" classes were further 3D classified with C2 symmetry for the closed conformation and without symmetry for the open and then the top class for each conformation were 3D refined to 4.5Å for the closed conformation, and 4.7Å for the

open conformation. A combined subset of the data was processed and refined to 4.2Å (Figure 5.9G) by including both conformations and applying C2 symmetry while masking out the extreme portion of the extended ends of the open conformation. Including both conformations as well as inducing symmetry on the open conformation may also introduce some distortions to the map.

The orientations contained within the tilted dataset provided much better coverage than the initial dataset (Figure 5.9C and D compared to A and B), and when the datasets were pooled, there was much greater coverage (Figure 5.9E and F). The CryoEF Efficiency number for the total combined dataset was 0.8 indicating that according to the parameters of the program, the pooled data should contain the minimum distribution of distributions to determine the structure. The model resulting from the pooled data set was generally much better, and secondary structure was beginning to emerge, with α helices clearly distinguishable in some portions of the map, especially throughout the center of the map. Despite this improvement *de novo* building of a model was limited still by the resolution of the map, as well as by the distortions due to preferred orientation. Chain tracing programs were utilized to attempt to trace a poly alanine chain through the map to establish the backbone including Pathwalker (Chen *et al.*, 2016), and Mainmast (Terashi and Kihara, 2018). Neither of these programs were able to produce a suitable model based on the current data set. Since the CD analysis of Clu1 indicated that it was largely α helical, and since the current map visibly contains a lot of α helices, a combination of α helix placement with Phenix and manually building in α helices in Coot were used to build a model into the map. This model could be utilized for molecular replacement phasing of crystal data from Clu1, or the map itself could also be used to help in future structural studies of Clu1.

### 5.3.3 X-Ray Crystallography

In addition to determining 3D structure of Clu1 by cryoEM X-Ray crystallography analysis of Clu1 has been undertaken. Based on experimentally determined and predicted features of Clu1Δ50 based on the characterization described thus far, selected conditions from

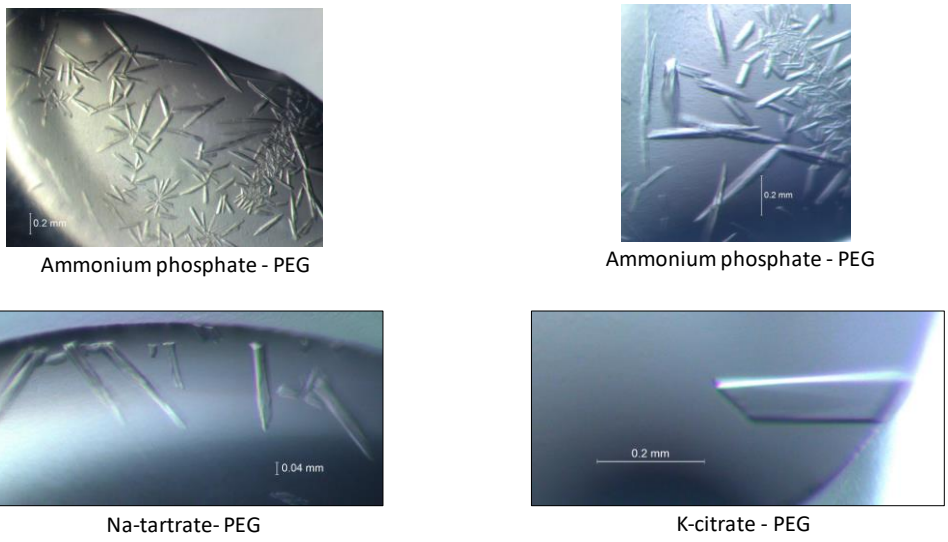


Figure 5.10 Clu1Δ50 crystals in different conditions.

Hampton Research PEG-ion screen were tested. Immediately several favorable conditions were found for crystals growth: 0.2 M ammonium phosphate-20% PEG 3350, 0.2 M potassium citrate-20% PEG 3350, 0.2 M lithium citrate-20% PEG 3350, 0.2 M K/Na-tartrate-PEG 20% 3350 and 0.2 M Na-tartrate-20% PEG 3350.

Crystals in those conditions grew at different protein concentration from 3.5 to 43 mg/ml and at different temperatures. In all conditions the crystals were of similar highly elongated shape as it is shown in Figure 5.10. The size of the crystals varied from 0.05 mm to 0.7 mm. The single crystals were a small fraction of total crystals with most of the crystals present in inseparable crystal clusters.

Based on the nature of the salts that were effective for crystal growth it was concluded that the major parameter promoting crystal growth was the nature of anions of the used salts. It appeared that anion in favorable salt must carry two or more charged groups (Figure 5.11). This hypothesis was tested experimentally. It appeared that crystals of the same shape were able to grow in 0.2 M potassium sulphate- 20% PEG 3350 and even 0.2 M EDTA- 20%PEG 3350. In all cases the shape of the crystals was very similar.

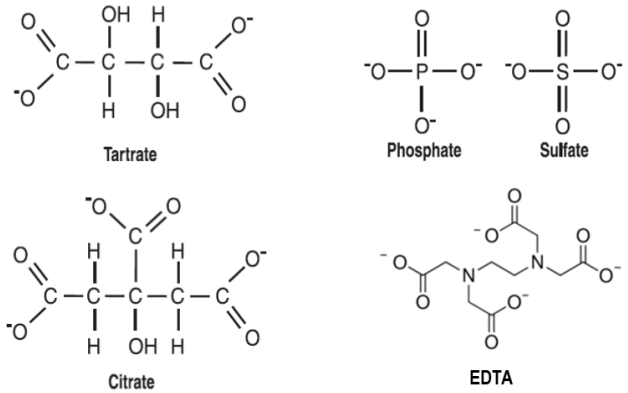


Figure 5.11. Chemical structure of anions favorable for Clu1ΔC50 crystal growth.

The crystals grew at different rates: crystals appeared overnight in ammonia phosphate-PEG buffer and took longer to grow in other conditions. They were allowed to mature for several days until no further visible growth detected. For X-Ray analysis crystals were soaked with cryoprotectant of different compositions. Working with the first crystals it was immediately clear that opening the crystallization wells with crystals in growing conditions resulted in very fast (few minutes) solubilization of the crystals back into solution. It was found that this can be avoided if the crystallization drop was first quickly mixed with a solution in which concentration of PEG was a little bit higher than in the drop (25%), and the salt concentration was about 50 mM rather than 200 mM in the crystallization conditions.

Mature crystals were soaked in cryoprotectants and flash frozen in liquid nitrogen. Initially glycerol was used as the main component of the cryoprotectant and the crystals were soaked in precipitant containing different concentrations of glycerol. The evaluation of X-Ray diffraction of the crystals and X-Ray data collection was done on Argonne National Laboratory Advanced Photon Source at the LS-CAT facility. The first crystals evaluated diffracted up to about 6 Å resolution with overall good shape of spots in cryoprotectant with about 20% glycerol.

Next, different cryoprotectants were used with different concentrations of glucose, sucrose, paraton, mineral oil, DMSO, etc. It was found that using 50 mM ammonium phosphate-20% PEG 3350-70% glucose resulted in the best diffraction with both, resolution and quality (shape and separation) of spots. Use of this cryoprotectant improved resolution to about 3.7 Å and therefore a several datasets were collected. However, diffraction from all examined crystals

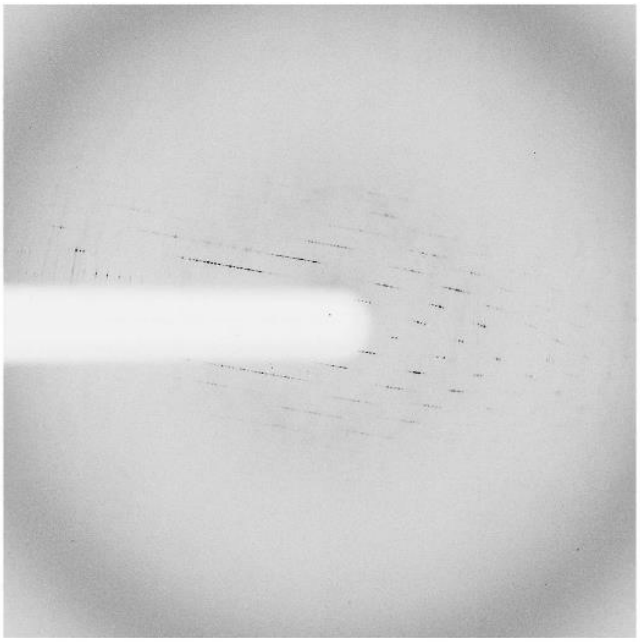

Figure 5.12. Representative frame from Clu1ΔC50 dataset. Crystal was grown in 0.2 M ammonium phosphate-20% PEG 3350

was highly anisotropic as it is shown in Figure 5.12 with the highest resolution only along major axis of ellipse formed by the spots. All crystals were tetragonal and of the same space group, I422. Unit cell dimensions from the best crystal were 88.895, 88.895, 660.045 Å. Calculated Matthew Coefficient (Vm) was 2.38 $Å^3$/Da which corresponds to 48.3% of solvent in the crystals. With obtained unit cell parameters and molecular mass of the Clu1ΔC50 protein of 139kDa there is one molecule in asymmetric unit (ASU) with 16 ASU in the unit cell. Many attempts have been made to get crystals of another shape, with lower degree of anisotropy and higher resolution by screening for different growth condition and use of different cryoprotectants, but none of these attempts have resulted in better data compared to the use of the ammonium phosphate-PEG-glucose combination so far.

Despite the anisotropic diffraction the collected datasets were processed by using autoPROC software in which staraniso server was also included. The autoPROC software produced several output files with two main output files: the first one, truncate.mtz, included

**Table 5.1. Statistics from Clu1Δ50 X-Ray Data Collection**

| | |
|---|---|
| Beamline | LS-CAT ID-G |
| Wavelength, (Å) | 0.97856 |
| Space group | I422 |
| Unit cell dimensions, *a*, *b*, *c*, (Å) | 88.9, 88.9, 660.0 |
| Angles, *a*, *b*, *g* (°) | 90.0, 90.0, 90.0 |
| Low resolution limit | 165.011 |
| High resolution limit | 3.676 |
| Total number of observations | 102911 |
| Total number unique | 4273 |
| Mean(I)/sd(I) | 11.0 |
| Completeness (spherical), (%) | 27.9 |
| Completeness (ellipsoidal), (%) | 87.5 |
| Multiplicity | 24.1 |
| CC(1/2) | 0.995 |
| No of molecules in asymmetric unit | 1 |
| Solvent fraction, (%) | 43 |
| Matthews coefficient, $Å^3$/Da | 2.38 |

reflection to the resolution at which the best statistics were obtained. In the best crystal this file included reflections up to 4.2 Å (Table 5.1). The autoPROC also produced output file with all reflections observed, alldata_unique.mtz. In the case of the best crystal this file included reflections up to 3.7 Å although statistics at the highest resolution was not satisfactory for the use in structure determination. In some other crystals reflections were registered up to 2.8 Å, with overall pure statistics.

## 5.4 Discussion

Overall, the structural analysis of Clu1 has revealed that it is an elongated dimer. CD analysis of Clu1 demonstrates that it is mostly composed of α helices, with minor contributions from β sheets, and the remainder from unstructured sections. Refolding assays demonstrate that $IP_6$ is required for the restoration of correct secondary structure of the protein after denaturation in urea and refolding. Negative stain electron microscopy analysis of these refolded particles demonstrates that in the presence of $IP_6$, Clu1 can refold into a conformation close to that of native protein.

Further characterization of Clu1 with cryo electron microscopy shows that Clu1 has two different conformations, a result which had previously been shown utilizing native mass spectrometry. Based on class averages, these two conformations differ in that one has a completely symmetrical overall structure, with each of the ends being in a “closed” conformation. The other has one “closed” end, and one “open” end. It is interesting that a class of particles with both “open” ends has never been observed. This may hold some significance to the biological function and regulation of Clu1.

In the process of collecting and analyzing data from Clu1, it became clear that preferred

orientation was an issue for further refinement of Clu1's structure using electron microscopy. In order to address this, data was collected with the microscope stage tilted with respect to the electron beam. By combining this data with previously collected data the resulting map was greatly improved, and a resolution of 4.2Å (based on estimates from Relion) was achieved. This map contained clear secondary structure, and while a full backbone was not able to be traced due to distortions from preferred orientations, a model containing α helices was constructed that may be useful for further structural studies of Clu1. X-Ray crystallography analysis of Clu1 was also conducted, and conditions for crystallization were determined, as well as cryoprotectant conditions. Several data sets were collected, and they all had the same I422 space group and highly elongated unit cell. Despite fairly good resolution in one dimension, data processing was hampered by the high degree on anisotropy.

To solve the crystal structure the phases must be determined. This could be done by a few different methods including molecular replacement, the use of heavy atoms derivatization and crystallization of selenomethionine containing protein. Traditional molecular replacement methods have proven to be difficult because there are no solved structures with high enough homology of Clu1. The highest homology that exists is about 16-18% and is found in the TPR domain of the Kinesin light chain. Traditional phasing with selenomethionine is done using protein expression in *E. coli*. To this end we cloned Clu1Δ50 in *E. coli* vector as GST-fusion protein (Vanderbilt pBG101 vector) and expressed in Rosetta *E.coli* strain. The Clu1Δ50 expressed in that strain/vector however all protein was in inclusion bodies. It was possible to produce NatA in *E. coli*, despite their lack of $IP_6$ due to the presence of ppGpp (Neubauer, 2012). This small negatively charge bacterial alarmone was found bound inside the $IP_6$ binding site of NatA and was sufficient for expression of the protein in the absence of $IP_6$. For Clu1, this does

not seem to be the case. There are a few published protocols for producing selenomethionine proteins in yeast (Malkowski *et al.*, 2007) but they are labor and material intensive and not used for proteins other than those presented in protocol developing publications. Use of heavy atom derivatization is a very challenging approach for any protein, and our attempts with using Os and Ni have not yet been successful.

Recently another approach to solving the phase problem has emerged. This approach is based on the use of cryoEM maps for phasing of crystallography data (Jackson and Terwilliger, 2015). This approach, however, has very specific requirements for the protein sequence and nature of the crystals. The protein must contain as many as possible structural domains with non-crystallographic symmetry (NSC) as well as very high solvent content. Although Clu1 contains 7 TPR repeats in each subunit of the dimer, it is not known whether they are organized a way that is suitable for this technique. The solvent content on Clu1Δ50 crystals is just 48.7%, which is lower than would be preferred for this method. Nevertheless, we attempted to utilize this method. The first steps of the protocol were completed, and all of the necessary working files were prepared. Using these EM maps with Phenix molecular replacement software several solutions were identified. However, the resulting electron density maps were of poor quality, and made little structural sense. With the quality of the current datasets the probability of getting wrong structure is too high, and further improvement of the X-Ray data set will be needed.

There are several future directions to determine the structure of Clu1. The first one is to continue screening for better crystallization/cryoprotectant conditions to find conditions for better crystals growth. The second direction is the crystallization of important functional and structural domains of the proteins, and subsequently to use them in structural model building and molecular replacement and refinement. Since the most distinctive features of Clu1 is the

presence of TPR repeats located in C-terminal part of protein, and several fragments of this portion of the protein have been selected to clone and express in *E. coli,* as well as in *S. cerevisiae*. From preliminary expression trials show that this fragment can be expressed independently. If this fragment can be expressed and purified, then determination of its structure, may provide a foothold for further structural characterization of Clu1.

# Chapter 6 Summary and Future Directions

## 6.1 Characterizing the Synthetic *E. coli* Inositide Pathway

The inositide signaling pathway is of great importance to cells. Various inositides are key for the regulation of many elements of Eukaryotic life, from marking various membranes within the inner membrane system of the cell, to signaling growth in response to extracellular factors, to mediating endocytosis. One difficulty encountered in studying inositide signaling is that the various lipids in the pathway are spatially separated among the various membranes of the cell. This spatial segregation of lipids from the enzymes that modulate them is key to the regulation of the pathway, but it can obfuscate the study of the actual activities of the enzymes. A system where there was no spatial segregation of the enzymes from their substrates would be a useful complement when interpreting the activities of the various kinases and phosphatases in the pathway. Another difficulty encountered in the study of the inositide pathway is the nature of studying the activity of enzymes that modify membrane lipids. Study of the enzymes *in vivo* provides the most accurate environment for the enzyme, however it is at times difficult to discern exactly what reactions are taking place in a living system full of other kinases and phosphatases that may be further modifying the products and substrates of the reaction in question. *In vitro* assays have long been used to establish the substrates that are acted on by inositide kinases and phosphatases, but the results obtained in these assays do not always reflect the activities of the enzymes in biological context. A system to study the activities of the inositide kinases in a biological context, that is devoid of endogenous inositide signaling enzymes would be of great benefit to understanding this key signaling pathway. To this end, we sought to develop a system for the expression of the key members of the inositide signaling pathway in *E. coli*, which do not

possess inositide signaling of their own.

Utilizing a duet vector system for the expression of combinations of genes in the phosphoinositide pathway we have reconstructed the basic components of phosphoinositide synthesis. With the expression of Pis1, we are able to demonstrate the formation of PI and when the PI kinases Pik1 or Vps34 are added, we see the conversion of PI to PI4P and PI3P respectively as expected. When the PIP kinase Mss4 is added to Pis1 and Pik1, we expectedly see the production of $PI(4,5)P_2$, however we also unexpectedly see the production of low levels of $PIP_3$. We confirmed the $PIP_3$ by expression of the $PIP_2$ kinase p110 and expectedly saw an increase in the production of $PIP_3$. The fact that in the absence of p110 we saw the production of $PIP_3$ indicated that this system might be able to reveal previously unobserved activities of phosphoinositide kinases. In addition to lipid inositide signaling we also reconstituted the soluble inositide signaling pathway by expressing PLC, which cleaves $PI(4,5)P_2$ and produces $I(1,4,5)P_3$, which when Ipk2 and Ipk1 are expressed can be sequentially phosphorylated to form $IP_6$.

One of the novel activities observed was when we co expressed Pis1, Vps34 and Mss4. As a PI4P 5 kinase, Mss4 would not be expected to have any activity with this combination of genes, however we observed a robust production of $PIP_3$, and the accumulation of an unidentified $PI(3,X)P_2$ intermediate. Especially of interest was the fact that all the genes utilized in this production of $PIP_3$ were genes from *S. cerevisiae*, where $PIP_3$ has not yet been identified indicating that *S. cerevisiae* may be capable of producing $PIP_3$, but regulation of the spatial localization of the kinases, the presence of phosphatases or some other factor may be preventing detectable levels of $PIP_3$.

## 6.2 Future directions

### 6.2.1 Further Define the Synthetic Pathway

While we here characterize the use of bacteria for the detection of unknown activities of inositide kinases, there is still room for further characterization. We observe the production of $PIP_3$ utilizing only *S. cerevisiae* genes, both in Pis1-Vps34-Mss4 expressing bacteria as well as in Pis1-Vps34-Fab1 expressing bacteria, however, we do not yet know exactly which enzymes are conducting what reactions. Further characterization of these bacterial strains may reveal more about the activities of the enzymes involved. By characterizing the intermediate $PI(3,X)P_2$s produced in each case, we may better understand the activities. This analysis can be conducted utilizing higher resolution HPLC (Sarkes and Rameh, 2010) to resolve the identities of the intermediate using $PI(3,4)P_2$ and $PI(3,5)P_2$ standards. Also utilizing treatment with wortmannin, an inhibitor of PI 3 kinases like Vps34 may to help characterize this activity. Identification of the PIP formed in Pis1 Fab1 bacteria to verify the direct synthesis of PI5P from PI by Fab1 is also of interest.

### 6.2.2 Expression of PI Kinase Enhancers

Another area for future development of the system would be in the characterization of genes that enhance the activity of the kinases. PIP kinases do not act in a vacuum, and in cells they are often members of large regulatory complexes, such asVps34 (Rostislavleva *et al.*, 2015) or Fig1Fab1Vac14 (Botelho *et al.*, 2008). The assembly of these complexes is important for maintenance of proper levels of PIPs, and their regulation is important for human health. Phosphatidylinositol Transfer Protein (PITP), are important regulators of PIPs, and in particular

Sec14 from *S. cerevisiae* is thought to enhance the activity of Pik1, which given the low activity of Pik1 compared to Vps34 in our system may also be of benefit simply from higher production of PI4P. Co expression of Pis1 Pik1 and the yeast PITP Sec14, compared to expression Pis1 and Pik1 alone and characterization of the resulting levels of PI4P, would be an important experiment for characterizing the activity of PI kinase enhancing proteins. Further expression of other enhancers of PI kinase activity would also be attractive utilizations of this system.

### 6.2.3 Expression of Phosphatases

A key feature of the lipid inositide signaling pathway is that it is regulated not only by kinases but also by a family of phosphatases. Expression of phosphatases in this system may reveal important details of their activity and regulation. Synaptojanin is a phosphatase that possesses both a 5 phosphatase domain as well as a Sac1 like phosphatase domain, and it is thought that upon binding to $PI(4,5)P_2$ Synaptojanin can sequentially remove the 5 phosphate, and then the 4 phosphate to produce PI, however the mechanism of this sequential reaction has not been worked out. There are also a variety of mutations that have been identified in humans that are associated with Alzheimer's disease, bipolar disorder and personality disorders in the 5-Phosphatase domain using PheWAS. Mutations associated with schizophrenia and psychosis have been identified in the Sac1 domain. The mechanisms for how these mutations are associated with disease, and how they impact the activity of the phosphatase have yet to be determined. Characterization of the impact that these disease associated mutations have on the processive activity of Synaptojanin's two phosphatase reactions in a clean system devoid of any other regulatory factors, but that still has a living membrane environment is an interesting application of this system. This characterization may lead to a better understanding of both the

mechanism of action of Synaptojanin, and the role that its disruption plays in human disease.

### 6.2.4 Expression of $IP_6$ Binding Proteins

As discussed in Chapter 2, there is a growing class of proteins that are bound to $IP_6$ as a structural cofactor. While some of these proteins can be expressed in *E. coli* due to some interesting interactions with negatively charged bacterial factors (Neubauer, 2012), it has also been demonstrated that others are unable to properly fold in the absence of $IP_6$ (Ring, O'Connell and Keegan, 2004). Since *E. coli* are a well characterized system for production of protein for structural biology, with well developed methods, tools and protocols for protein expression and purification, a method for the *E. coli* production of these $IP_6$ binding proteins could be an important tool for the characterization of these proteins. By expressing these $IP_6$ binding proteins along with the genes required for production of $IP_6$, it may be possible to express proteins in *E. coli* that otherwise would be unable to fold. This would allow for further characterization of this growing class of proteins.

### 6.2.5 Optimize and Integrate Operon

In addition to utilizing sets of plasmids for expression of phosphoinositide kinases, we also demonstrate that a synthetic operon can be utilized. Further development of this operon-based system for expression of the inositol pathway would be of benefit to the system. Inclusion of the inositol transporter IolT to the operon, as well as development of alternative operons composed of different sets of genes, like the set required for $PIP_3$ production, would greatly add to the potential uses of the operon. Additionally, in bacteria such as *B. subtilis*, there are already operons for the degradation of inositol, and the regulatory elements from this operon are

dependent on levels of inositol in the environment (Yoshida *et al.*, 1997). Putting the regulatory elements from this operon into the synthetic phosphoinositide operon would allow for the induction of the synthetic phosphoinositide signaling pathway by adding inositol to the media, further simplifying the control of the system. Additionally, integration of the synthetic operon into the *E. coli* genome, eliminating the need for plasmid-based expression would also greatly enhance the utility of the system.

### 6.3 Characterization of the $IP_6$ Binding Protein Clu1

Inositol Phosphates have long been studied for their role as signaling molecules, specifically with the identification of $IP_3$ as a second messenger involved in the release of calcium. It was later discovered that $IP_3$ can be sequentially phosphorylated to produce a whole array of IPs. In the study of these more highly phosphorylated IPs, it became clear that some of these IPs do not behave in ways that are typical of second messengers, leading to the hypothesis that they may be acting with a different mechanism from $IP_3$. With the solving of the crystal structures of several proteins that have $IP_6$ bound within solvent inaccessible pockets with very high affinity, it became clear that in some cases $IP_6$ could bind to proteins as a structural cofactor. The York lab then developed a screen to identify such proteins, and found two proteins NatA, and Clu1, which both bind to $IP_6$ via interactions with TPR domains (Pham, 2012). This motif of $IP_6$ binding to the center of a supersolenoid composed of tandem repeat domains has been identified in a growing family of proteins. Further characterization of $IP_6$ binding proteins is needed to better understand the role that $IP_6$ is playing in these proteins.

Clu1 is a protein that is known to be important for the regulation of mitochondrial distribution, as well as binding to mRNAs for proteins involved in oxidative phosphorylation.

Here we demonstrate that the C-terminus of Clu1 is required for binding to RNA. We also demonstrate that $IP_6$ binding, but not RNA binding is required for Clu1's role in mitochondrial distribution. We show that Clu1 is a dimeric protein, and that there are two different conformations of the Clu1 dimer. We also show that $IP_6$ may play a role in the stability of the Clu1 dimer, and that it is required for in vitro refolding of the protein. We also utilized single particle Cryo EM to characterize Clu1. This structural characterization was hampered by limitations of preferred orientation of Clu1 in vitrified ice. We have produced a map of Clu1, that while partially deformed due to preferred orientations, is below 5Å, and may prove useful for the future characterization of Clu1. We have also undertaken X-Ray crystallography characterization of Clu1. We have determined conditions for Clu1 crystallization, begun optimization of cryo conditions for the crystals and collected diffraction data from these crystals. The data collected is highly anisotropic, and has been challenging to process, however the overall quality of the data is promising.

## 6.4 Future Directions

### 6.4.1 Further Characterization of Clu1 with Cryo Electron Microscopy

One approach to future characterization of Clu1 with CryoEM would be to try to find a way to disrupt the preferred orientations that are currently a problem for determining the structure of Clu1. One way to do this may be to try a different method of grid preparation. It is possible that further screening of additives and surfactants may identify conditions that result in a better distribution of orientations. Another possibility is the use of a carbon support film, since it has been observed during initial screening of conditions that Clu1 particles had a high affinity for the carbon support. Since the Clu1 dimer is smaller than 500kDa, it is likely that use of a

graphene oxide support film would be better than a more traditional carbon support film, due to its better electron transparency. Both commercially prepared graphene support films, as well as preparation of homemade graphene support films at the Vanderbilt CryoEM facility are options for this approach. Additionally, as alternative grid prep approaches, like the development of the Spotiton (Dandey *et al.*, 2018), become more routine, use of such techniques may also be useful to address the problems with Clu1 sample preparation. Other approaches to be utilized include changing the construct used for preparation of grids. Negative stain grids were analyzed utilizing Clu1 with the MBP fusion protein still attached, and the presence of the MBP tag could serve to alter the preferred orientation of Clu1. By collecting a dataset of Clu1 with MBP, a dataset with better distribution of orientations could be attained. In addition to trying to change the orientations, another approach would be to collect a large enough data set to include a number of rare orientations, and then weight these rare orientations more heavily during data processing. Additionally, collection of more data with an even greater tilt could help to deal with the preferred orientations. Both of these approaches would greatly benefit from collection of data with a Titan Krios microscope.

### 6.4.2 Further Characterization of Clu1 with X-ray Crystallography

In addition to Cryo Electron microscopy, X-Ray crystallography has been utilized to study Clu1. While so far this approach has not yielded a solved structure, there are several avenues for future work in this direction. Further screening of crystallization conditions, such as utilizing additive screens have been undertaken, as well as further screening for better cryo conditions. Additionally, there is a growing family of protein structures that have been crystallized and solved with an MBP fusion protein tag left on the protein (Smyth *et al.*, 2003;

Waugh, 2016). For these proteins crystallization was aided by the presence of the MBP tag, and for a subset of the proteins the phases were solved with molecular replacement using the structure of MBP as the search model. Such an approach may be able to aid in the solving of Clu1's structure, and to this end screens for crystallization conditions for Clu1 with its MBP fusion protein tag have been undertaken. Also, currently underway is the development of expression vectors for expression of various truncations of Clu1 that include the important $IP_6$ binding TPR domain of Clu1, both in *E. coli* and in *S. cerevisiae*. Additionally, expression of Clu1 in *E. coli* that also produce $IP_6$ is an attractive direction to produce protein for structural characterization. In order to solve the phasing problem there are also several approaches being considered. One of these is to utilize the current CryoEM map, as well as future, less distorted maps to phase the X-Ray diffraction data. Another approach is to utilize methods of incorporation of selenomethionine in *S. cerevisiae* or other eukaryotic expression systems like *K. lactis*. Also, Molecular Replacement based on the small fragments of Clu1 that have homology to already solved proteins, and on poly alanine traces generated from the CryoEM maps are being pursued. While the structure of Clu1 remains elusive, with continued work from a variety of approaches, we are confident that this goal is within reach.

# BIBLIOGRAPHY


Alcázar-Román, A. R. and Wente, S. R. (2008) 'Inositol polyphosphates: A new frontier for regulating gene expression', *Chromosoma*. doi: 10.1007/s00412-007-0126-4.

Ando, T. *et al.* (2016) 'The host protein CLUH participates in the subnuclear transport of influenza virus ribonucleoprotein complexes', *Nature Microbiology*. Nature Publishing Group, 1(8). doi: 10.1038/nmicrobiol.2016.62.

Arnone, A. and Perutz, M. F. (1974) 'Structure of inositol hexaphosphate-human deoxyhaemoglobin complex', *Nature*. doi: 10.1038/249034a0.

Ballou, C. E. and Lee, Y. C. (1964) 'The Structure of a Myoinositol Mannoside from Mycobacterium Tuberculosis Glycolipid', *Biochemistry*. doi: 10.1021/bi00893a014.

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

Barker, C. J. *et al.* (2004) 'Complex changes in cellular inositol phosphate complement accompany transit through the cell cycle', *Biochemical Journal*. doi: 10.1042/BJ20031872.

Berridge, M. J. *et al.* (1983) 'Changes in the levels of inositol phosphates after agonist-dependent hydrolysis of membrane phosphoinositides', *Biochemical Journal*, 212(2), pp. 473–482. doi: 10.1042/bj2120473.

Bettaney, K. E. *et al.* (2013) 'A systematic approach to the amplified expression, functional characterization and purification of inositol transporters from Bacillus subtilis.', *Molecular membrane biology*. doi: 10.3109/09687688.2012.729093.

Blind, R. D. *et al.* (2014) 'The signaling phospholipid PIP $_3$ creates a new interaction surface on the nuclear receptor SF-1', *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1416740111.

Bonangelino, C. J. *et al.* (2002) 'Osmotic stress-induced increase of phosphatidylinositol 3,5-bisphosphate requires Vac14p, an activator of the lipid kinase Fab1p', *Journal of Cell Biology*. doi: 10.1083/jcb.200201002.

Boronenkov, I. V. and Anderson, R. A. (1995) 'The sequence of phosphatidylinositol-4-phosphate 5-kinase defines a novel family of lipid kinases', *Journal of Biological Chemistry*. doi: 10.1074/jbc.270.7.2881.

Botelho, R. J. *et al.* (2008) 'Assembly of a Fab1 phosphoinositide kinase signaling complex requires the Fig4 phosphoinositide phosphatase', *Molecular Biology of the Cell*. doi: 10.1091/mbc.E08-04-0405.

Brear, P. *et al.* (2016) 'Specific inhibition of CK2α from an anchor outside the active site', *Chemical Science*. doi: 10.1039/c6sc02335e.

Brown, J. R. and Auger, K. R. (2011) 'Phylogenomics of phosphoinositide lipid kinases: Perspectives on the evolution of second messenger signaling and drug discovery', *BMC Evolutionary Biology*. doi: 10.1186/1471-2148-11-4.

Bruno, S. *et al.* (2000) 'Hemoglobin Crystals', pp. 683–692.

Cantley, L. C. (2002) 'The phosphoinositide 3-kinase pathway', *Science*. doi: 10.1126/science.296.5573.1655.

Chalhoub, N. and Baker, S. J. (2009) 'PTEN and the PI3-Kinase Pathway in Cancer', *Annual Review of Pathology: Mechanisms of Disease*. doi: 10.1146/annurev.pathol.4.110807.092311.

Charenton, C., Wilkinson, M. E. and Nagai, K. (2019) 'Mechanism of 5' splice site transfer for human spliceosome activation.', *Science (New York, N.Y.)*. United States, 364(6438), pp. 362–367. doi: 10.1126/science.aax3289.

Chen, K.-E. E. *et al.* (2018) 'Molecular Basis for Membrane Recruitment by the PX and C2 Domains of Class II Phosphoinositide 3-Kinase-C2α', *Structure*. United States: Elsevier Ltd., 26(12), pp. 1612-1625.e4. doi: 10.1016/j.str.2018.08.010.

Chen, M. *et al.* (2016) 'De Novo modeling in cryo-EM density maps with Pathwalking', *Journal of Structural Biology*. doi: 10.1016/j.jsb.2016.06.004.

Chen, Q. *et al.* (2017) 'Structural basis of arrestin-3 activation and signaling', *Nature Communications*. England, 8(1), p. 1427. doi: 10.1038/s41467-017-01218-8.

Chishti, A. H. *et al.* (1998) 'The FERM domain: A unique module involved in the linkage of cytoplasmic proteins to the membrane', *Trends in Biochemical Sciences*. doi: 10.1016/S0968-0004(98)01237-7.

Chow, C. Y. *et al.* (2007) 'Mutation of FIG4 causes neurodegeneration in the pale tremor mouse and patients with CMT4J', *Nature*. doi: 10.1038/nature05876.

Chow, C. Y. *et al.* (2009) 'Deleterious Variants of FIG4, a Phosphoinositide Phosphatase, in Patients with ALS', *American Journal of Human Genetics*. doi: 10.1016/j.ajhg.2008.12.010.

Clark, J. *et al.* (2011) 'Quantification of PtdInsP 3 molecular species in cells and tissues by mass spectrometry', *Nature Methods*. doi: 10.1038/nmeth.1564.

Clarke, B. P. *et al.* (2019) 'A synthetic biological approach to reconstitution of inositide signaling pathways in bacteria', *Advances in Biological Regulation*. doi: 10.1016/j.jbior.2019.100637.

Clarke, J. H., Wang, M. and Irvine, R. F. (2010) 'Localization, regulation and function of Type II phosphatidylinositol 5-phosphate 4-kinases', *Advances in Enzyme Regulation*. doi:

10.1016/j.advenzreg.2009.10.006.

Coates, M. L. (1975) ‘Hemoglobin function in the vertebrates: An evolutionary model’, *Journal of Molecular Evolution*. doi: 10.1007/BF01794636.

Collins, B. M. *et al.* (2002) ‘Molecular architecture and functional model of the endocytic AP2 complex’, *Cell*, 109(4), pp. 523–535. doi: 10.1016/S0092-8674(02)00735-3.

Cox, R. T. and Spradling, A. C. (2009) ‘Clueless, a conserved Drosophila gene required for mitochondrial subcellular localization, interacts genetically with parkin’, *DMM Disease Models and Mechanisms*, 2(9–10), pp. 490–499. doi: 10.1242/dmm.002378.

D’Andrea, L. D. and Regan, L. (2003) ‘TPR proteins: The versatile helix’, *Trends in Biochemical Sciences*. doi: 10.1016/j.tibs.2003.10.007.

Dandey, V. P. *et al.* (2018) ‘Spotiton: New features and applications’, *Journal of Structural Biology*. doi: 10.1016/j.jsb.2018.01.002.

Delgado-Ramírez, M. *et al.* (2018) ‘Regulation of Kv2.1 channel inactivation by phosphatidylinositol 4,5-bisphosphate’, *Scientific Reports*. doi: 10.1038/s41598-018-20280-w.

Deng, S. *et al.* (2019) ‘Structure and Mechanism of Acetylation by the N-Terminal Dual Enzyme NatA/Naa50 Complex.’, *Structure (London, England : 1993)*. United States. doi: 10.1016/j.str.2019.04.014.

Desrivières, S. *et al.* (1998) ‘MSS4, a phosphatidylinositol-4-phosphate 5-kinase required for organization of the actin cytoskeleton in Saccharomyces cerevisiae’, *Journal of Biological Chemistry*. doi: 10.1074/jbc.273.25.15787.

Dick, R. A. *et al.* (2018) ‘Inositol phosphates are assembly co-factors for HIV-1.’, *Nature*. England. doi: 10.1038/s41586-018-0396-4.

Dimmer, K. S. *et al.* (2002) ‘Genetic basis of mitochondrial function and morphology in Saccharomyces cerevisiae’, *Molecular Biology of the Cell*. doi: 10.1091/mbc.01-12-0588.

Dove, S. K. *et al.* (2002) ‘Vac14 controls PtdIns(3,5)P2 synthesis and Fab1-dependent protein trafficking to the multivesicular body’, *Current Biology*. doi: 10.1016/S0960-9822(02)00891-6.

El-Gebali, S. *et al.* (2019) ‘The Pfam protein families database in 2019’, *Nucleic Acids Research*. doi: 10.1093/nar/gky995.

Emerling, B. M. *et al.* (2013) ‘Depletion of a putatively druggable class of phosphatidylinositol kinases inhibits growth of p53-Null tumors’, *Cell*, 155(4), p. 844. doi: 10.1016/j.cell.2013.09.057.

Erdman, S. *et al.* (1998) ‘Pheromone-regulated genes required for yeast mating differentiation’, *Journal of Cell Biology*. doi: 10.1083/jcb.140.3.461.

Fedele, C. G. *et al.* (2010) 'Inositol polyphosphate 4-phosphatase II regulates PI3K/Akt signaling and is lost in human basal-like breast cancers', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1015245107.

Fica, S. M. *et al.* (2017) 'Structure of a spliceosome remodelled for exon ligation', *Nature*. doi: 10.1038/nature21078.

Fica, S. M. *et al.* (2019) 'A human postcatalytic spliceosome structure reveals essential roles of metazoan factors for exon ligation', *Science*. doi: 10.1126/science.aaw5569.

Fields, S. D., Conrad, M. N. and Clarke, M. (1998) 'The S. cerevisiae CLU1 and D. discoideum cluA genes are functional homologues that influence mitochondrial morphology and distribution', *Journal of cell science*, 111 ( Pt 1, pp. 1717–1727. Available at: http://jcs.biologists.org/cgi/reprint/111/12/1717.

Flanagan, C. A. *et al.* (1993) 'Phosphatidylinositol 4-kinase: Gene structure and requirement for yeast cell viability', *Science*. doi: 10.1126/science.8248783.

Folch, J. (1942) 'Brain Cephalin, a Mixture of Phosphatides. Separation From It of Phosphatidyl Serine, Phosphatidyl Ethanolamine, and a Fraction Containing an Inositol Phosphatide', *Journal of Biological Chemistry*, 146(1), pp. 35–44.

Ford, M. G. J. *et al.* (2001) 'Simultaneous binding of PtdIns (4,5) P 2 and clathrin by AP180 in the nucleation of clathrin lattices on membranes', *Science*, 291(5506), pp. 1051–1055. doi: 10.1126/science.291.5506.1051.

Franco-Echevarría, E. *et al.* (2017) 'The crystal structure of mammalian inositol 1,3,4,5,6-pentakisphosphate 2-kinase reveals a new zinc-binding site and key features for protein function', *Journal of Biological Chemistry*, 292(25), pp. 10534–10548. doi: 10.1074/jbc.M117.780395.

Frederick, J. P. *et al.* (2005) 'An essential role for an inositol polyphosphate multikinase, Ipk2, in mouse embryogenesis and second messenger production', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0503706102.

Fridy, P. C. *et al.* (2007) 'Cloning and characterization of two human VIP1-like inositol hexakisphosphate and diphosphoinositol pentakisphosphate kinases', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M704656200.

Fruman, D. A. *et al.* (1999) 'Impaired B cell development and proliferation in absence of phosphoinositide 3-kinase p85α', *Science*. doi: 10.1126/science.283.5400.393.

Fruman, D. A. *et al.* (2017) 'The PI3K Pathway in Human Disease', *Cell*. doi: 10.1016/j.cell.2017.07.029.

Fruman, D. a, Meyers, R. E. and Cantley, L. C. (1998) 'Phosphoinositide kinases.', *Annual review of biochemistry*. doi: 10.1146/annurev.biochem.67.1.481.

De Fusco, C. *et al.* (2017) ‘A fragment-based approach leading to the discovery of a novel binding site and the selective CK2 inhibitor CAM4066.’, *Bioorganic & medicinal chemistry*. England, 25(13), pp. 3471–3482. doi: 10.1016/j.bmc.2017.04.037.

Gaidarov, I. and Keen, J. H. (1999) ‘Phosphoinositide-AP-2 interactions required for targeting to plasma membrane clathrin-coated pits’, *Journal of Cell Biology*. doi: 10.1083/jcb.146.4.755.

Van Galen, J. *et al.* (2012) ‘Interaction of GAPR-1 with lipid bilayers is regulated by alternative homodimerization’, *Biochimica et Biophysica Acta - Biomembranes*. doi: 10.1016/j.bbamem.2012.04.016.

Gao, J. *et al.* (2014) ‘CLUH regulates mitochondrial biogenesis by binding mRNAs of nuclear-encoded mitochondrial proteins’, *Journal of Cell Biology*, 207(2), pp. 213–223. doi: 10.1083/jcb.201403129.

Gary, J. D. *et al.* (1998) ‘Fab1p is essential for PtdIns(3)P 5-kinase activity and the maintenance of vacuolar size and membrane homeostasis’, *Journal of Cell Biology*. doi: 10.1083/jcb.143.1.65.

González, B. *et al.* (2010) ‘Inositol 1,3,4,5,6-pentakisphosphate 2-kinase is a distant IPK member with a singular inositide binding site for axial 2-OH recognition’, *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0912979107.

Gordon, V. M. and Leppla, S. H. (1994) ‘Proteolytic activation of bacterial toxins: Role of bacterial and host cell proteases’, *Infection and Immunity*.

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

Gosein, V. and Miller, G. J. (2013) ‘Conformational stability of inositol 1,3,4,5,6-pentakisphosphate 2-kinase (IPK1) dictates its substrate selectivity.’, *The Journal of biological chemistry*. United States, 288(52), pp. 36788–36795. doi: 10.1074/jbc.M113.512731.

Gottlieb, L. and Marmorstein, R. (2018) ‘Structure of Human NatA and Its Regulation by the Huntingtin Interacting Protein HYPK.’, *Structure (London, England : 1993)*. United States. doi: 10.1016/j.str.2018.04.003.

Greiner, R., Konietzny, U. and Jany, K. D. (1993) ‘Purification and characterization of two phytases from *Escherichia coli*’, *Archives of biochemistry and biophysics*. doi: 10.1006/abbi.1993.1261.

Gruninger, R. J. *et al.* (2012) ‘Substrate binding in protein-tyrosine phosphatase-like inositol polyphosphatases’, *Journal of Biological Chemistry*, 287(13), pp. 9722–9730. doi: 10.1074/jbc.M111.309872.

Guo, S. *et al.* (1999) ‘SAC1-like domains of yeast SAC1, INP52, and INP53 and of human synaptojanin encode polyphosphoinositide phosphatases’, *Journal of Biological Chemistry*. doi: 10.1074/jbc.274.19.12990.

Halstead, J. R. *et al.* (2001) 'A novel pathway of cellular phosphatidylinositol(3,4,5)-trisphosphate synthesis is regulated by oxidative stress', *Current Biology*. doi: 10.1016/S0960-9822(01)00121-X.

Hama, H. *et al.* (1999) 'Direct involvement of phosphatidylinositol 4-phosphate in secretion in the yeast Saccharomyces cerevisiae', *Journal of Biological Chemistry*. doi: 10.1074/jbc.274.48.34294.

Han, G. S. *et al.* (2002) 'The Saccharomyces cerevisiae LSB6 gene encodes phosphatidylinositol 4-kinase activity', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M207996200.

Haselbach, D. *et al.* (2018) 'Structure and Conformational Dynamics of the Human Spliceosomal Bact Complex', *Cell*, 172(3), pp. 454-464.e11. doi: 10.1016/j.cell.2018.01.010.

Hatch, A. J. and York, J. D. (2010) 'SnapShot: Inositol phosphates', *Cell*. Elsevier, 143(6), pp. 1030-1030.e1. doi: 10.1016/j.cell.2010.11.045.

Hayashi, K. I. *et al.* (2008) 'Small-molecule agonists and antagonists of F-box protein-substrate interactions in auxin perception and signaling', *Proceedings of the National Academy of Sciences of the United States of America*, 105(14), pp. 5632–5637. doi: 10.1073/pnas.0711146105.

Hendricks, K. B. *et al.* (1999) 'Yeast homologue of neuronal frequenin is a regulator of phosphatidylinosrtol-4-OH kinase', *Nature Cell Biology*. doi: 10.1038/12058.

Heymont, J. *et al.* (2000) 'TEP1, the yeast homolog of the human tumor suppressor gene PTEN/MMAC1/TEP1, is linked to the phosphatidylinositol pathway and plays a role in the developmental process of sporulation', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.97.23.12672.

Hinnebusch, A. G. (2006) 'eIF3: a versatile scaffold for translation initiation complexes', *Trends in Biochemical Sciences*. doi: 10.1016/j.tibs.2006.08.005.

Hollander, M. C., Blumenthal, G. M. and Dennis, P. A. (2011) 'PTEN loss in the continuum of common cancers, rare syndromes and mouse models', *Nature Reviews Cancer*. doi: 10.1038/nrc3037.

Iegre, J. *et al.* (2018) 'Second-generation CK2α inhibitors targeting the αd pocket', *Chemical Science*. doi: 10.1039/c7sc05122k.

Irvine, R. F. *et al.* (1986) 'The inositol tris/tetrakisphosphate pathway - Demonstration of Ins(l,4,5)P3 3-kinase activity in animal tissues', *Nature*. doi: 10.1038/320631a0.

Irvine, R. F. (2016) 'A short history of inositol lipids', *Journal of Lipid Research*. doi: 10.1194/jlr.r071712.

Ito, H. *et al.* (1983) 'Transformation of intact yeast cells treated with alkali cations', *Journal of Bacteriology*.

Ivetac, I. *et al.* (2009) 'Regulation of PI(3)K/Akt signalling and cellular transformation by inositol polyphosphate 4-phosphatase-1', *EMBO Reports*. doi: 10.1038/embor.2009.28.

Jackson, R. N. and Terwilliger, T. C. (2015) 'X-ray structure determination using low-resolution electron microscopy maps for molecular replacement', (November 2017). doi: 10.1038/nprot.2015.069.

Jacques, D. A. *et al.* (2016) 'HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis', *Nature*. doi: 10.1038/nature19098.

Joung, M.-J., Mohan, S. K. and Yu, C. (2012) 'Molecular level interaction of inositol hexaphosphate with the C2B domain of human synaptotagmin I.', *Biochemistry*. United States, 51(17), pp. 3675–3683. doi: 10.1021/bi300005w.

Kantner, M., Schöll, E. and Yanchuk, S. (2014) 'Delay-induced patterns in a two-dimensional lattice of coupled oscillators', *Scientific Reports*. doi: 10.1038/srep08522.

Kavanaugh, J. S., Rogers, P. H. and Arnone, A. (2005) 'Crystallographic evidence for a new ensemble of ligand-induced allosteric transitions in hemoglobin: The T-to-THigh quaternary transitions', *Biochemistry*, 44(16), pp. 6101–6121. doi: 10.1021/bi047813a.

Keegan, L. P., Gallo, A. and O'Connell, M. A. (2001) 'The many roles of an RNA editor', *Nature Reviews Genetics*. doi: 10.1038/35098584.

Kelly, B. T. *et al.* (2014) 'Clathrin adaptors. AP2 controls clathrin polymerization with a membrane-activated switch.', *Science (New York, N.Y.)*. United States, 345(6195), pp. 459–463. doi: 10.1126/science.1254836.

Kim, D. *et al.* (2013) 'TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions', *Genome Biology*. doi: 10.1186/gb-2013-14-4-r36.

Kim, Y. J., Guzman-Hernandez, M. L. and Balla, T. (2011) 'A highly dynamic ER-derived phosphatidylinositol-synthesizing organelle supplies phosphoinositides to cellular membranes', *Developmental Cell*. doi: 10.1016/j.devcel.2011.09.005.

de la Rosa-Trevín, J. M. *et al.* (2016) 'Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy', *Journal of Structural Biology*. doi: 10.1016/j.jsb.2016.04.010.

Labriola, J., Zhou, Y. and Nagar, B. (2018) 'Structural analysis of the bacterial effector, AvrA, identifies a critical helix involved in MKK4-substrate recognition.', *Biochemistry*. United States. doi: 10.1021/acs.biochem.8b00512.

Larkin, R. M. *et al.* (2016) 'REDUCED CHLOROPLAST COVERAGE genes from Arabidopsis thaliana help to establish the size of the chloroplast compartment', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1515741113.

Lee, W.-K. *et al.* (2013) 'Structural and functional insights into the regulation mechanism of

CK2 by IP6 and the intrinsically disordered protein Nopp140', *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1304670110.

Lee, Y. R., Chen, M. and Pandolfi, P. P. (2018) 'The functions and regulation of the PTEN tumour suppressor: new modes and prospects', *Nature Reviews Molecular Cell Biology*. doi: 10.1038/s41580-018-0015-0.

Lee, Y. S. *et al.* (2007) 'Regulation of a cyclin-CDK-CDK inhibitor complex by inositol pyrophosphates', *Science*. doi: 10.1126/science.1139080.

Lennartz, F. *et al.* (2017) 'Structure-Guided Identification of a Family of Dual Receptor-Binding PfEMP1 that Is Associated with Cerebral Malaria', *Cell Host and Microbe*. United States: Elsevier Inc., 21(3), pp. 403–414. doi: 10.1016/j.chom.2017.02.009.

Letunic, I. and Bork, P. (2018) '20 years of the SMART protein domain annotation resource', *Nucleic Acids Research*. doi: 10.1093/nar/gkx922.

Li, X. *et al.* (2013) 'Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM', *Nature Methods*. doi: 10.1038/nmeth.2472.

Lim, D. *et al.* (2000) 'Crystal structures of Escherichia coli phytase and its complex with phytate', *Nature Structural Biology*, 7(2), pp. 108–113. doi: 10.1038/72371.

Lin, D. H. *et al.* (2018) 'Structural and functional analysis of mRNA export regulation by the nuclear pore complex', *Nature Communications*. England: Springer US, 9(1), pp. 1–19. doi: 10.1038/s41467-018-04459-3.

Liu, S. *et al.* (2017) 'Structure of the yeast spliceosomal postcatalytic P complex', *Science*. United States, 358(6368), pp. 1278–1283. doi: 10.1126/science.aar3462.

Löfke, C. *et al.* (2008) 'Alternative metabolic fates of phosphatidylinositol produced by phosphatidylinositol synthase isoforms in Arabidopsis thaliana', *Biochemical Journal*. doi: 10.1042/BJ20071371.

Logan, D. C., Scott, L. and Tobin, A. K. (2003) 'The genetic control of plant mitochondrial morphology and dynamics', *Plant Journal*, 36(4), pp. 500–509. doi: 10.1046/j.1365-313X.2003.01894.x.

Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*. doi: 10.1186/s13059-014-0550-8.

Luisi, B. *et al.* (1990) 'Structure of deoxy-quaternary haemoglobin with liganded β subunits', *Journal of Molecular Biology*, 214(1), pp. 7–14. doi: 10.1016/0022-2836(90)90139-D.

Lupardus, P. J. *et al.* (2008) 'Small Molecule – Induced Allosteric Cysteine Protease Domain', *Science*, 322(October), pp. 265–268. doi: 10.1126/science.1162403.

Macbeth, M. R., Schubert, H. L., VanDemark, A. F., Lingam, A. T., Hill, C. P., Bass, B. L., *et al.*

(2005) ‘Structural biology: Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing’, *Science*. doi: 10.1126/science.1113150.

Macbeth, M. R., Schubert, H. L., VanDemark, A. F., Lingam, A. T., Hill, C. P. and Bass, B. L. (2005) ‘Structural biology: Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing’, *Science*. doi: 10.1126/science.1113150.

Malkowski, M. G. *et al.* (2007) ‘Blocking S-adenosylmethionine synthesis in yeast allows selenomethionine incorporation and multiwavelength anomalous dispersion phasing’, *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0610337104.

Mallery, D. L. *et al.* (2018) ‘IP6 is an HIV pocket factor that prevents capsid collapse and promotes DNA synthesis’, *eLife*, 7. doi: 10.7554/eLife.35335.

Mani, M. *et al.* (2007) ‘The Dual Phosphatase Activity of Synaptojanin1 Is Required for Both Efficient Synaptic Vesicle Endocytosis and Reavailability at Nerve Terminals’, *Neuron*. doi: 10.1016/j.neuron.2007.10.032.

Manning, G. *et al.* (2002) ‘Evolution of protein kinase signaling from yeast to man’, *Trends in Biochemical Sciences*. doi: 10.1016/S0968-0004(02)02179-5.

Marsh, D. J. *et al.* (1998) ‘Mutation spectrum and genotype-phenotype analyses in Cowden disease and Bannayan-Zonana syndrome, two hamartoma syndromes with germline PTEN mutation’, *Human Molecular Genetics*. doi: 10.1093/hmg/7.3.507.

Matthews, M. M. *et al.* (2016) ‘Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity.’, *Nature structural & molecular biology*. United States, 23(5), pp. 426–433. doi: 10.1038/nsmb.3203.

McConnachie, G. *et al.* (2003) ‘Interfacial kinetic analysis of the tumour suppressor phosphatase, PTEN: Evidence for activation by anionic phospholipids’, *Biochemical Journal*. doi: 10.1042/BJ20021848.

Meijer, H. J. G. *et al.* (2001) ‘Identification of a new polyphosphoinositide in plants, phosphatidylinositol 5-monophosphate (PtdIns5P), and its accumulation upon osmotic stress’, *Biochemical Journal*. doi: 10.1042/0264-6021:3600491.

Michell, R. H. (2008) ‘Inositol derivatives: Evolution and functions’, *Nature Reviews Molecular Cell Biology*. doi: 10.1038/nrm2334.

Milano, S. K. *et al.* (2006) ‘Nonvisual arrestin oligomerization and cellular localization are regulated by inositol hexakisphosphate binding’, *Journal of Biological Chemistry*, 281(14), pp. 9812–9823. doi: 10.1074/jbc.M512703200.

Miller, S. *et al.* (2010) ‘Shaping development of autophagy inhibitors with the structure of the lipid kinase Vps34’, *Science*. doi: 10.1126/science.1184429.

Mitra, P. *et al.* (2004) ‘A novel phosphatidylinositol(3,4,5)P3pathway in fission yeast’, *Journal of Cell Biology*. doi: 10.1083/jcb.200404150.

Monserrate, J. P. and York, J. D. (2010) ‘Inositol phosphate synthesis and the nuclear processes they affect’, *Current Opinion in Cell Biology*. doi: 10.1016/j.ceb.2010.03.006.

Monteleone, L. R. *et al.* (2019) ‘A Bump-Hole Approach for Directed RNA Editing’, *Cell Chemical Biology*. United States: Elsevier Ltd., 26(2), pp. 269-277.e5. doi: 10.1016/j.chembiol.2018.10.025.

Montpetit, B. *et al.* (2011) ‘A conserved mechanism of DEAD-box ATPase activation by nucleoporins and InsP6 in mRNA export’, *Nature*. Nature Publishing Group, 472(7342), pp. 238–244. doi: 10.1038/nature09862.

Moravcevic, K. *et al.* (2015) ‘Comparison of Saccharomyces cerevisiae F-BAR domain structures reveals a conserved inositol phosphate binding site’, *Structure*. United States: Elsevier, 23(2), pp. 352–363. doi: 10.1016/j.str.2014.12.009.

Morii, H. *et al.* (2010) ‘A revised biosynthetic pathway for phosphatidylinositol in Mycobacteria’, *Journal of Biochemistry*. doi: 10.1093/jb/mvq093.

Morita, Y. S. *et al.* (2010) ‘Stress-induced synthesis of phosphatidylinositol 3-phosphate in mycobacteria’, *Journal of Biological Chemistry*. doi: 10.1074/jbc.M110.119263.

Mulugu, S. *et al.* (2007) ‘A conserved family of enzymes that phosphorylate inositol hexakisphosphate’, *Science*. doi: 10.1126/science.1139099.

Mumberg, D., Müller, R. and Funk, M. (1995) ‘Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds’, *Gene*. doi: 10.1016/0378-1119(95)00037-7.

Naydenova, K. and Russo, C. J. (2017) ‘Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy’, *Nature Communications*. doi: 10.1038/s41467-017-00782-3.

Neubauer, J. (2012) *The N--terminal Acetyltransferase NatA Binds Inositol Hexakisphosphate And Exhibits Conformational Flexibility*. Duke University.

Neuwald, A. F. *et al.* (1992) ‘cysQ, a gene needed for cysteine synthesis in Escherichia coli K-12 only during aerobic growth’, *Journal of Bacteriology*. doi: 10.1128/jb.174.2.415-425.1992.

Niebuhr, K. *et al.* (2002) ‘Conversion of PtdIns(4, 5)P2 into PtdIns(5)P by the S.flexneri effector IpgD reorganizes host cell morphology’, *EMBO Journal*. doi: 10.1093/emboj/cdf522.

Nikawa, J., Kodaki, T. and Yamashita, S. (1988) ‘Expression of the Saccharomyces cerevisiae PIS gene and synthesis of phosphatidylinositol in Escherichia coli.’, *Journal of bacteriology*, 170(10), pp. 4727–4731. doi: 10.1128/jb.170.10.4727-4731.1988.

Odorizzi, G., Babst, M. and Emr, S. D. (1998) 'Fab1p Ptdlns(3)P 5-kinase function essential for protein sorting in the multivesicular body', *Cell*. doi: 10.1016/S0092-8674(00)81707-9.

Otto, J. C. *et al.* (2007) 'Alterations in an inositol phosphate code through synergistic activation of a G protein and inositol phosphate kinases.', *Proceedings of the National Academy of Sciences USA*. doi: 10.1073/pnas.0705729104.

Otto, J. C. and York, J. D. (2010) 'Molecular manipulation and analysis of inositol phosphate and pyrophosphate levels in Mammalian cells.', *Methods in molecular biology (Clifton, N.J.)*. doi: 10.1007/978-1-60327-175-2_3.

Ouyang, Z. *et al.* (2016a) 'Structural Basis and IP6 Requirement for Pds5-Dependent Cohesin Dynamics', *Molecular Cell*. doi: 10.1016/j.molcel.2016.02.033.

Ouyang, Z. *et al.* (2016b) 'Structural Basis and IP6Requirement for Pds5-Dependent Cohesin Dynamics', *Molecular Cell*. doi: 10.1016/j.molcel.2016.02.033.

Paoli, M. *et al.* (1996) 'Crystal structure of T state haemoglobin with oxygen bound at all four haems', *Journal of Molecular Biology*, 256(4), pp. 775–792. doi: 10.1006/jmbi.1996.0124.

Pham, T. (2012) *A Role for Inositol Hexakisphosphate in N-terminal Acetylation and Mitochondrial Distribution*. Duke University.

Pittet, D. *et al.* (1989) 'Mass changes in inositol tetrakis- and pentakisphosphate isomers induced by chemotactic peptide stimulation in HL-60 cells', *Journal of Biological Chemistry*.

Prochazkova, K. *et al.* (2009) 'Structural and molecular mechanism for autoprocessing of MARTX toxin of vibrio cholerae at multiple sites', *Journal of Biological Chemistry*, 284(39), pp. 26557–26568. doi: 10.1074/jbc.M109.025510.

Pruitt, R. N. *et al.* (2009) 'Structure-function analysis of inositol hexakisphosphate-induced autoprocessing in Clostridium difficile toxin A', *Journal of Biological Chemistry*, 284(33), pp. 21934–21940. doi: 10.1074/jbc.M109.018929.

Puri, A. W. *et al.* (2010) 'Rational design of inhibitors and activity-based probes targeting clostridium difficile virulence factor TcdB', *Chemistry and Biology*. Elsevier Ltd, 17(11), pp. 1201–1211. doi: 10.1016/j.chembiol.2010.09.011.

Rameh, L. E. *et al.* (1997) 'A new pathway for synthesis of phosphatidylinositol-4,5-bisphosphate', *Nature*. doi: 10.1038/36621.

Raucher, D. *et al.* (2000) 'Phosphatidylinositol 4,5-bisphosphate functions as a second messenger that regulates cytoskeleton-plasma membrane adhesion', *Cell*. doi: 10.1016/S0092-8674(00)81560-3.

Reddy, N. R., Sathe, S. K. and Salunkhe, D. K. (1982) 'Phytates in legumes and cereals', *Advances in Food Research*. doi: 10.1016/S0065-2628(08)60110-X.

Rigden, D. J. *et al.* (1999) 'Polyanionic inhibitors of phosphoglycerate mutase: Combined structural and biochemical analysis', *Journal of Molecular Biology*, 289(4), pp. 691–699. doi: 10.1006/jmbi.1999.2848.

Ring, G. M., O'Connell, M. A. and Keegan, L. P. (2004) 'Purification and assay of recombinant ADAR proteins expressed in the yeast Pichia pastoris or in Escherichia coli.', *Methods in molecular biology (Clifton, N.J.)*.

Rostislavleva, K. *et al.* (2015) 'Structure and flexibility of the endosomal Vps34 complex reveals the basis of its function on membranes', *Science*. doi: 10.1126/science.aac7365.

Sablin, E. P. *et al.* (2015) 'Structure of Liver Receptor Homolog-1 (NR5A2) with PIP3hormone bound in the ligand binding pocket', *Journal of Structural Biology*. doi: 10.1016/j.jsb.2015.09.012.

Saiardi, A. *et al.* (1999) 'Synthesis of diphosphoinositol pentakisphosphate by a newly identified family of higher inositol polyphosphate kinases', *Current Biology*. doi: 10.1016/S0960-9822(00)80055-X.

Sarkes, D. and Rameh, L. E. (2010) 'A novel HPLC-based approach makes possible the spatial characterization of cellular PtdIns5P and other phosphoinositides', *Biochemical Journal*. doi: 10.1042/BJ20100129.

Schatton, D. *et al.* (2017) 'CLUH regulates mitochondrial metabolism by controlling translation and decay of target mRNAs', *The Journal of cell biology*, 216(3), pp. 675–693. doi: 10.1083/jcb.201607019.

Schatton, D. and Rugarli, E. I. (2018) 'A concert of RNA-binding proteins coordinates mitochondrial function', *Critical Reviews in Biochemistry and Molecular Biology*. Taylor & Francis, 53(6), pp. 652–666. doi: 10.1080/10409238.2018.1553927.

Scheres, S. H. W. (2012) 'RELION: Implementation of a Bayesian approach to cryo-EM structure determination', *Journal of Structural Biology*. doi: 10.1016/j.jsb.2012.09.006.

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

Schu, P. V. *et al.* (1993) 'Phosphatidylinositol 3-kinase encoded by yeast VPS34 gene essential for protein sorting', *Science*. doi: 10.1126/science.8385367.

Sen, A. *et al.* (2015) 'Clueless, a protein required for mitochondrial function, interacts with the PINK1-Parkin complex in Drosophila', *DMM Disease Models and Mechanisms*, 8(6), pp. 577–589. doi: 10.1242/dmm.019208.

Sheard, L. B. *et al.* (2010) 'Jasmonate perception by inositol-phosphate-potentiated COI1-JAZ co-receptor', *Nature*. doi: 10.1038/nature09430.

Shears, S. B. (2009) 'Molecular basis for the integration of inositol phosphate signaling

pathways via human ITPK1', *Advances in Enzyme Regulation*. doi: 10.1016/j.advenzreg.2008.12.008.

Shen, A. *et al.* (2009) 'Mechanistic and structural insights into the proteolytic activation of Vibrio cholerae MARTX toxin', *Nature Chemical Biology*, 5(7), pp. 469–478. doi: 10.1038/nchembio.178.

Shen, A. *et al.* (2011) 'Defining an allosteric circuit in the cysteine protease domain of Clostridium difficile toxins', *Nature Structural and Molecular Biology*. Nature Publishing Group, 18(3), pp. 364–371. doi: 10.1038/nsmb.1990.

Shisheva, A. (2008) 'PIKfyve: Partners, significance, debates and paradoxes', *Cell Biology International*. doi: 10.1016/j.cellbi.2008.01.006.

SMITH, G. N., WORREL, C. S. and SWANSON, A. L. (1949) 'Inhibition of bacterial esterases by chloramphenicol.', *Journal of bacteriology*.

Smyth, D. R. *et al.* (2003) 'Crystal structures of fusion proteins with large-affinity tags', *Protein Science*. doi: 10.1110/ps.0243403.

Snoek, T., Verstrepen, K. J. and Voordeckers, K. (2016) 'How do yeast cells become tolerant to high ethanol concentrations?', *Current Genetics*. doi: 10.1007/s00294-015-0561-3.

Solomon, O. *et al.* (2013) 'Global regulation of alternative splicing by adenosine deaminase acting on RNA (ADAR)', *RNA*. doi: 10.1261/rna.038042.112.

Song, M. S., Salmena, L. and Pandolfi, P. P. (2012) 'The functions and regulation of the PTEN tumour suppressor', *Nature Reviews Molecular Cell Biology*. doi: 10.1038/nrm3330.

Sorzano, C. O. S. *et al.* (2004) 'XMIPP: A new generation of an open-source image processing package for electron microscopy', *Journal of Structural Biology*. doi: 10.1016/j.jsb.2004.06.006.

Spiegelberg, B. D. *et al.* (1999) 'Cloning and characterization of a mammalian lithium-sensitive bisphosphate 3'-nucleotidase inhibited by inositol 1,4-bisphosphate', *Journal of Biological Chemistry*. doi: 10.1074/jbc.274.19.13619.

Stack, J. H. *et al.* (1993) 'A membrane-associated complex containing the Vps15 protein kinase and the Vps34 PI 3-kinase is essential for protein sorting to the yeast lysosome-like vacuole.', *The EMBO journal*. doi: 10.1111/hisn.12004_8.

Stack, J. H. *et al.* (1995) 'Vesicle-mediated protein transport: Regulatory interactions between the Vps15 protein kinase and the Vps34 PtdIns 3-kinase essential for protein sorting to the vacuole in yeast', *Journal of Cell Biology*. doi: 10.1083/jcb.129.2.321.

Stack, J. H. and Emr, S. D. (1994) 'Vps34p required for yeast vacuolar protein sorting is a multiple specificity kinase that exhibits both protein kinase and phosphatidylinositol-specific PI 3-kinase activities', *Journal of Biological Chemistry*.

Stevenson-Paulik, J. *et al.* (2006) 'Inositol phosphate metabolomics: Merging genetic perturbation with modernized radiolabeling methods', *Methods*. doi: 10.1016/j.ymeth.2006.05.012.

Stevenson, J. M. *et al.* (2000) 'Inositol signaling and plant growth', *Trends in Plant Science*. doi: 10.1016/S1360-1385(00)01652-6.

Stolz, L. E. *et al.* (1998) 'INP51, a yeast inositol polyphosphate 5-phosphatase required for phosphatidylinositol 4,5-bisphosphate homeostasis and whose absence confers a cold-resistant phenotype', *Journal of Biological Chemistry*. doi: 10.1074/jbc.273.19.11852.

Storey, D. J. *et al.* (1984) 'Stepwise enzymatic dephosphorylation of inositol 1,4,5-trisphosphate to inositol in liver', *Nature*. doi: 10.1038/312374a0.

Streb, H. *et al.* (1983) 'Release of Ca2+ from a nonmitochondrial intracellular store in pancreatic acinar cells by inositol-1,4,5-trisphosphate', *Nature*, 306(5938), pp. 67–69. doi: 10.1038/306067a0.

Sun, J. and MacKinnon, R. (2017) 'Cryo-EM Structure of a KCNQ1/CaM Complex Reveals Insights into Congenital Long QT Syndrome', *Cell*. doi: 10.1016/j.cell.2017.05.019.

Švančarová, P. and Betáková, T. (2018) 'Conserved methionine 165 of matrix protein contributes to the nuclear import and is essential for influenza A virus replication', *Virology Journal*. doi: 10.1186/s12985-018-1056-x.

Tan, X. *et al.* (2007) 'Mechanism of auxin perception by the TIR1 ubiquitin ligase', *Nature*. doi: 10.1038/nature05731.

Tang, G. *et al.* (2007) 'EMAN2: An extensible image processing suite for electron microscopy', *Journal of Structural Biology*. doi: 10.1016/j.jsb.2006.05.009.

Terashi, G. and Kihara, D. (2018) 'De novo main-chain modeling for em maps using MAINMAST', *Nature Communications*. doi: 10.1038/s41467-018-04053-7.

Terebiznik, M. R. *et al.* (2002) 'Elimination of host cell Ptdlns(4, 5)P2 by bacterial SigD promotes membrane fission during invasion by Salmonella', *Nature Cell Biology*. doi: 10.1038/ncb854.

Thorsell, A. G. *et al.* (2009) 'Crystal structure of human diphosphoinositol phosphatase 1', *Proteins: Structure, Function and Bioinformatics*, 77(1), pp. 242–246. doi: 10.1002/prot.22489.

Tolia, N. H. and Joshua-Tor, L. (2006) 'Strategies for protein coexpression in Escherichia coli', *Nature Methods*. doi: 10.1038/nmeth0106-55.

Toste Rêgo, A. and da Fonseca, P. C. A. (2019) 'Characterization of Fully Recombinant Human 20S and 20S-PA200 Proteasome Complexes', *Molecular Cell*, pp. 1–10. doi: 10.1016/j.molcel.2019.07.014.

Viaud, J. *et al.* (2014) 'Phosphatidylinositol 5-phosphate: A nuclear stress lipid and a tuner of membranes and cytoskeleton dynamics', *BioEssays*. doi: 10.1002/bies.201300132.

Viaud, J. *et al.* (2016) 'Phosphoinositides: Important lipids in the coordination of cell dynamics', *Biochimie*. doi: 10.1016/j.biochi.2015.09.005.

Vlach, J. *et al.* (2018) 'Structural basis for targeting avian sarcoma virus Gag polyprotein to the plasma membrane for virus assembly.', *The Journal of biological chemistry*. United States. doi: 10.1074/jbc.RA118.003944.

Volpicelli-Daley, L. A. *et al.* (2010) 'Phosphatidylinositol-4-phosphate 5-kinases and phosphatidylinositol 4,5-bisphosphate synthesis in the brain', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M110.132191.

Vornlocher, H. P. *et al.* (1999) 'A 110-kilodalton subunit of translation initiation factor eIF3 and an associated 135-kilodalton protein are encoded by the Saccharomyces cerevisiae TIF32 and TIF31 genes', *Journal of Biological Chemistry*. doi: 10.1074/jbc.274.24.16802.

Wakim, J. *et al.* (2017) 'CLUH couples mitochondrial distribution to the energetic and metabolic status', *Journal of Cell Science*, 130(11), pp. 1940–1951. doi: 10.1242/jcs.201616.

Walch-Solimena, C. and Novick, P. (1999) 'The yeast phosphatidylinositol-4-OH kinase Pik1 regulates secretion at the Golgi', *Nature Cell Biology*. doi: 10.1038/70319.

Waller, D. A. and Liddington, R. C. (1990) 'Refinement of a partially oxygenated T state human haemoglobin at 1.5 Å resolution', *Acta Crystallographica Section B*. doi: 10.1107/S0108768190000313.

Wallroth, A. and Haucke, V. (2018) 'Phosphoinositide conversion in endocytosis and the endolysosomal system', *Journal of Biological Chemistry*. doi: 10.1074/jbc.R117.000629.

Walsh, J. P., Caldwell, K. K. and Majerus, P. W. (1991) 'Formation of phosphatidylinositol 3-phosphate by isomerization from phosphatidylinositol 4-phosphate.', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.88.20.9184.

Wan, R. *et al.* (2017) 'Structure of an Intron Lariat Spliceosome from Saccharomyces cerevisiae', *Cell*. doi: 10.1016/j.cell.2017.08.029.

Wan, R. *et al.* (2019) 'Structures of the Catalytically Activated Yeast Spliceosome Reveal the Mechanism of Branching', *Cell*. doi: 10.1016/j.cell.2019.02.006.

Wang, H. *et al.* (2012) 'Structural basis for an inositol pyrophosphate kinase surmounting phosphate crowding', *Nature Chemical Biology*. Nature Publishing Group, 8(1), pp. 111–116. doi: 10.1038/nchembio.733.

Wang, H. *et al.* (2014) 'IP6K structure and the molecular determinants of catalytic specificity in an inositol phosphate kinase family', *Nature Communications*. England: Nature Publishing Group, 5(May), pp. 1–12. doi: 10.1038/ncomms5178.

Wang, Q. *et al.* (2015) 'Autoinhibition of Bruton's tyrosine kinase (Btk) and activation by soluble inositol hexakisphosphate', *eLife*. England, 4, pp. 1–31. doi: 10.7554/elife.06074.

Wang, Z. H., Clark, C. and Geisbrecht, E. R. (2016) 'Drosophila clueless is involved in Parkin-dependent mitophagy by promoting VCP-mediated Marf degradation', *Human Molecular Genetics*, 25(10), pp. 1946–1964. doi: 10.1093/hmg/ddw067.

Watson, P. J. *et al.* (2016) 'Insights into the activation mechanism of class i HDAC complexes by inositol phosphates', *Nature Communications*. doi: 10.1038/ncomms11262.

Waugh, D. S. (2016) 'Crystal structures of MBP fusion proteins', *Protein Science*. doi: 10.1002/pro.2863.

Westermann, B. and Neupert, W. (2000) 'Mitochondria-targeted green fluorescent proteins: Convenient tools for the study of organelle biogenesis in Saccharomyces cerevisiae', *Yeast*. doi: 10.1002/1097-0061(200011)16:15<1421::AID-YEA624>3.0.CO;2-U.

Whitaker, M. (1985) 'Polyphosphoinositide hydrolysis is associated with exocytosis in adrenal medullary cells', *FEBS Letters*. doi: 10.1016/0014-5793(85)80858-9.

Whitaker, M. and Aitchison, M. (1985) 'Calcium-dependent polyphosphoinositide hydrolysis is associated with exocytosis in vitro', *FEBS Letters*. doi: 10.1016/0014-5793(85)81167-4.

Whitfield, H. *et al.* (2018) 'A Fluorescent Probe Identifies Active Site Ligands of Inositol Pentakisphosphate 2-Kinase', *Journal of Medicinal Chemistry*. doi: 10.1021/acs.jmedchem.8b01022.

Whitman, M. *et al.* (1988) 'Type I phosphatidylinositol kinase makes a novel inositol phospholipid, phosphatidylinositol-3-phosphate', *Nature*. doi: 10.1038/332644a0.

Wiedemann, C., Bellstedt, P. and Görlach, M. (2013) 'CAPITO - A web server-based analysis and plotting tool for circular dichroism data', *Bioinformatics*. doi: 10.1093/bioinformatics/btt278.

Wild, R. *et al.* (2016) 'Control of eukaryotic phosphate homeostasis by inositol polyphosphate sensor domains', *Science*. doi: 10.1126/science.aad9858.

Wilkinson, M. E. *et al.* (2017) 'Postcatalytic spliceosome structure reveals mechanism of 3′–splice site selection', *Science*. doi: 10.1126/science.aar3729.

Williams, S. P., Gillaspy, G. E. and Perera, I. Y. (2015) 'Biosynthesis and possible functions of inositol pyrophosphates in plants', *Frontiers in Plant Science*. doi: 10.3389/fpls.2015.00067.

Wilson, M. S. C. *et al.* (2015) 'A novel method for the purification of inositol phosphates from biological samples reveals that no phytate is present in human plasma or urine', *Open Biology*. doi: 10.1098/rsob.150014.

Wilson, M. S. C., Livermore, T. M. and Saiardi, A. (2013) 'Inositol pyrophosphates: between

signalling and metabolism', *Biochemical Journal*. doi: 10.1042/BJ20130118.

Wrobel, A. G. *et al.* (2019) 'Temporal Ordering in Endocytic Clathrin-Coated Vesicle Formation via AP2 Phosphorylation', *Developmental Cell*, 50(4), pp. 494-508.e11. doi: 10.1016/j.devcel.2019.07.017.

York, J. D. *et al.* (1999) 'A phospholipase C-dependent inositol polyphosphate kinase pathway required for efficient messenger RNA export', *Science*, 285(5424), pp. 96–100. doi: 10.1126/science.285.5424.96.

Yoshida, K. I. *et al.* (1997) 'Organization and transcription of the myo-inositol operon, iol, of Bacillus subtilis', *Journal of Bacteriology*. doi: 10.1128/jb.179.14.4591-4598.1997.

Yoshida, Soshi *et al.* (1994) 'A novel gene, STT4, encodes a phosphatidylinositol 4-kinase in the PKC1 protein kinase pathway of Saccharomyces cerevisiae', *Journal of Biological Chemistry*.

Yoshida, Satoshi *et al.* (1994) 'Genetic interactions among genes involved in the STT4-PKC1 pathway of Saccharomyces cerevisiae', *MGG Molecular & General Genetics*. doi: 10.1007/BF00283416.

Zhan, X. *et al.* (2018a) 'Structure of a human catalytic step I spliceosome', *Science*. doi: 10.1126/science.aar6401.

Zhan, X. *et al.* (2018b) 'Structures of the human pre-catalytic spliceosome and its precursor spliceosome.', *Cell research*. England. doi: 10.1038/s41422-018-0094-7.

Zhang, K. (2016) 'Gctf: Real-time CTF determination and correction', *Journal of Structural Biology*. doi: 10.1016/j.jsb.2015.11.003.

Zhang, X. *et al.* (1997) 'Phosphatidylinositol-4-phosphate 5-kinase isozymes catalyze the synthesis of 3-phosphate-containing phosphatidylinositol signaling molecules', *Journal of Biological Chemistry*. doi: 10.1074/jbc.272.28.17756.

Zhang, X. *et al.* (2017) 'An Atomic Structure of the Human Spliceosome', *Cell*. doi: 10.1016/j.cell.2017.04.033.

Zhang, X. *et al.* (2019) 'Structures of the human spliceosomes before and after release of the ligated exon', *Cell Research*. doi: 10.1038/s41422-019-0143-x.

Zhang, Z.-M. *et al.* (2017) 'Mechanism of host substrate acetylation by a YopJ family effector.', *Nature plants*. England, 3, p. 17115. doi: 10.1038/nplants.2017.115.

Zhang, Z. M. *et al.* (2016) 'Structure of a pathogen effector reveals the enzymatic mechanism of a novel acetyltransferase family.', *Nature structural & molecular biology*. United States: Nature Publishing Group, 23(9), pp. 847–852. doi: 10.1038/nsmb.3279.

Zhu, Q. *et al.* (1997) 'The cluA- mutant of Dictyostelium identifies a novel class of proteins required for dispersion of mitochondria', *Proceedings of the National Academy of Sciences of the*

*United States of America*, 94(14), pp. 7308–7313. doi: 10.1073/pnas.94.14.7308.

Zoncu, R. *et al.* (2007) 'Loss of endocytic clathrin-coated pits upon acute depletion of phosphatidylinositol 4,5-bisphosphate', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0611733104.