FUNCTION OF THE PIT-ACCESSORY PROTEIN

CHARACTERIZING THE FUNCTION OF THE PIT-ACCESSORY PROTEIN (PAP) IN *SINORHIZOBIUM MELILOTI*

By DANIEL HSIEH, B.Sc. MOLECULAR BIOLOGY AND GENETICS

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements for the

Degree of Master of Science

Master of Science (2017)

McMaster University (Biology), Hamilton, Ontario, Canada

Title: Characterizing the function of the Pit-accessory protein (Pap) in *Sinorhizobium meliloti*

Author: Daniel Hsieh, B.Sc. Biology (McMaster University)

Supervisor: Dr. Turlough M. Finan

Number of Pages: i-xv, 1-119

Lay Abstract

Microbes acquire and assimilate phosphorus (P) in the form of inorganic phosphate (Pi) through a variety of mechanisms. Pit (Pi transporter) are a family of diverse transporters found in all kingdoms of life. Unlike other Pit systems, the *Sinorhizobium meliloti pit* gene is encoded in an operon with a protein of unknown function, denoted the *pit*-accessory protein (*pap*). Using *S. meliloti* Pap-Pit and orthologues from other bacteria as model systems, we demonstrate that Pap functions as a positive modulator of Pi uptake via Pit, as Pap is required for active uptake of Pi. Pap-Pit systems are found in 30% of all bacteria and archaea, and thus are broadly distributed. Understanding the mechanism of Pap-Pit has biotechnological applications, as multiple Pap-Pit systems are present in phosphorus-accumulating bacteria utilized for waste-water treatment.

Abstract

Microorganisms primarily acquire phosphorus (P) in the form of inorganic phosphate (PO_4^{-3} or Pi) through expression of a suite of phosphate scavenging or phosphate transporter systems in response to growth-limiting environmental phosphate. One such system is the Pit family of single protein Pi transport systems found in all domains. These vary in size from 300 to 800 amino acids (a.a.) in size.

Previously, the *pit* gene of the soil bacterium *Sinorhizobium meliloti*, was found to encode a 334 a.a. Pi uptake system (K_M 1-2μM) that is repressed in low Pi conditions. However, the *S. meliloti pit* gene is encoded in an operon and overlaps the coding sequence of a protein of unknown function, which was denoted as *pap* (pit-accessory protein).

Using a conditional Pi-transport deficient mutant strain of *S. meliloti*, the effects of *pap* or *pit* mutations on Pit-mediated Pi uptake were studied by conducting growth experiments in minimal media (with Pi as the sole source of P) and Pi uptake experiments. Both *pap* and *pit* deletions resulted in a loss of growth and Pi uptake, which could be complemented by integration of the *pap* and *pit* genes into the deletion locus.

Heterologous Pap-Pit systems from *Bacteroides thetaiotaomicron* and *Shewanella oneidensis* were found to have K_M values (17 and 8.5 μM, respectively) similar to previously reported values of *S. meliloti* Pap-Pit. However, the *Shewanella* Pit protein was capable of transporting Pi in the absence of the cognate Pap protein, albeit with greatly reduced velocity at all measured concentrations.

Pap-Pit orthologs were identified in ~2000 diverse prokaryotic proteomes using Pfam motifs of Pit (PHO4) and Pap (PhoU_div) protein domains. *pap-pit* operons were found in a third of all proteomes, and were predicted to be a co-transcribed operon in >95% of cases. This provided additional evidence that Pap is directly involved in Pit-mediated Pi uptake, and also that Pap-Pit systems have a significant role in microbial Pi uptake.

Pap protein sequences and structures show striking similarities with that of PhoU, a protein of unknown function implicated as a modulator of the Pst uptake system. Pap and PhoU proteins share highly conserved putative metal-binding motifs (E/DXXXD) of which several Pap missense mutations were found to result in reduced Pi transport. This suggests that like PhoU, Pap may function as a modulator of Pi uptake by an interaction with its cognate transporter, Pit. However, the molecular mechanisms of PhoU and Pap proteins have yet to be defined.

Acknowledgements

First and foremost, I would like to thank Dr. Turlough Finan for the opportunity to study under his mentorship, for it would not have been possible without the patience and guidance. I would also like to thank Turlough for the chance to work in his lab as a summer work-study and fourth-year thesis student, which sparked my excitement of microbiology and genetics.

I am also grateful for the all the work done by previous members of the Finan Lab which made this research possible. In particular, Dr. George diCenzo who took the time out of his busy days to answer my questions and concerns but also as someone to exchange ideas with. I would also like to thank Drs. Rahat Zaheer, Dr. Vahid Husseininaveh and Hui Zhao, of which many of their strains and plasmid constructs were used in my study and the results achieved here would not be possible without their hard-work. And a thank you to the Lauren Tiller and Zahed Muhammed, who helped out with cloning experiments which saved me a lot of time during the last few months of my thesis. And a special thank you to Michael MacLeod and Zahed Muhammed for keeping me company during the long days in the lab this past year.

And lastly, I would like to thank my family and friends for their unwavering support during the last two years. It would not have been possible without the contributions of all these people.

Table of Contents

List of Figures and Tables

List of Abbreviations and Symbols

DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
P	Phosphorus
Pi	Inorganic Phosphate
SmPapPit	*S. meliloti* Pap-Pit
EcPitA	*E. coli* PitA
BtPapPit	*Bacteroides thetaiotaomicron* Pap-Pit
SoPapPit	*Shewanella oneidensis* Pap-Pit
Δ	Genetic deletion
Φ	Bacteriophage transduction
MFS	Major Facilitator Superfamily

Declaration of Academic Achievement

I declare that the research contribution which follows is original work, completed and written by me, with editorial assistance from my supervisor Dr. Turlough M. Finan.

The following individuals contributed to experimentation:

- In Chapter 2.5 – "Previous work in *S. meliloti pap-pit*", the deletion constructs of the *pap-pit* genes were generated by Dr. Rahat Zaheer and Hui Zhao. The Pap missense mutations were originally constructed by Hui Zhao and Dr. Vahid Husseininaveh and were used in this experiment.
- In Chapter 2.6 – "Conditionally phosphate uptake-deficient mutant of *S. meliloti*", the background strain was originally constructed by Dr. George C. diCenzo as part of his publication entitled "PhoU allows rapid adaptation to high phosphate concentrations by modulating PstSCAB transport rate in *Sinorhizobium meliloti*" J Bacteriol JB-00143.
- In Chapter 2.7.2 – "Cloning of heterologous *pap-pit* systems in pMW119", the cloning experimental was conducted by Lauren Tiller.

Chapter 1 - Literature Review and Introduction

1.1 - Phosphorous acquisition in prokaryotic organisms

Phosphorus is an essential element to life and is a major component of vital biomolecules such as nucleic acids (DNA/RNA), membrane phospholipids, and a component of signal transduction processes through phosphorylation. Microorganisms primarily acquire phosphorus in the form of inorganic phosphate (PO_4^{-3} or Pi). Many microorganisms express a suite of phosphate scavenging and high affinity phosphate uptake systems in response to low concentrations of environmental phosphate (Hsieh and Wanner, 2010).

1.1.1 - PhoR-PhoB and the Pho Regulon

To respond to changing environmental levels of phosphate, bacteria and archaea employ a two-component system consisting of a histidine-kinase sensor (commonly denoted as PhoR) and cognate transcription factor (commonly denoted as PhoB) (Makino *et al*., 1989). The N-terminal domain of PhoR has two membrane spanning helices and the protein dimerizes in response to low concentrations of Pi and autophosphorylates the C-terminal histidine phosphorylation domain on the intracellular (cytoplasmic) side of the protein (Figure 1-1) (Makino *et al.,* 1989). Once dimerized, PhoR phosphorylates PhoB, which dimerizes and binds to a specific sequence called the Pho Box motif in the promoter regions of various genes (Makino *et al.*, 1996). These genes encode a suite of phosphate starvation proteins, including alkaline phosphatases, transporters of carbon-

phosphate molecules (such as phosphonates), membrane phospholipids replacement, and high-affinity transporters of Pi (van Veen, 1997; Geiger *et al.*, 1999; Yuan *et al.*, 2006).

Mutations that alter the regulation of the Pho Regulon not only inhibit the response to low Pi but also have pleiotropic effects that significantly reduce or attenuate bacterial pathogenicity (Crépin *et al.*, 2011). For the nitrogen-fixing symbiotic bacterium *Sinorhizobium meliloti* (*S. meliloti*), the ability to transport phosphates has been shown to be required for nitrogen fixing symbiosis with alfalfa (Oresnik *et al.*, 1994; Bardin *et al.*, 1996; Yuan *et al.*, 2006).

1.1.2 - PstSCAB uptake systems

The Pi-specific transporter (Pst) system is an ATP-binding cassette (ABC)-type transporter which consists of three component proteins: PstS, the periplasmic phosphate-specific solute binding protein; PstC and PstA, the transmembrane heterodimeric permease responsible for the translocation of Pi across the membrane; and PstB, the ATPase that hydrolyzes ATP to generate the energy required for Pi uptake (Cox *et al.*, 1989; Chan and Torriani, 1996). PstSCAB systems are high-affinity systems with a K_M in the low micromolar range, between 0.1 - 1μM Pi. Interestingly, PhoU is also proposed to function as a modulator of Pi uptake via PstSCAB (Rice *et al.*, 2009; Lubin *et al.*, 2016; diCenzo *et al.*, 2017).

1.1.3 - PhnCDE(T) transport systems

The PhnCDE(T) phosphonate transport (Phn) system is an ABC-type transporter permitting uptake of phosphonates (molecules with C-P bonds) that is induced under low Pi conditions in a PhoB-dependent manner (Metcalf and Wanner, 1993; Yakovleva *et al*., 1998). The *E. coli* PhnCDE system differs from the *S. meliloti* PhnCDET, as in *S. meliloti* the translocase component consists of a PhnET protein heterodimer. However, the Phn systems in *S. meliloti* and *Mycobacterium smegmatis* were observed to function as Pi uptake systems (Voegele *et al.,* 1997; Gebhard *et al.,* 2006).

1.1.4 - Pit transport systems

The Pit (Pi-transport) family of single protein Pi uptake systems is found in all domains of life. These proteins vary in size from 300 to 800 amino acids (a.a.) (Bøttger and Pedersen, 2011). Pit proteins are members of the Major Facilitator Superfamily (MFS), a large group of proteins that transport a broad variety of small molecules, including inorganic and organic molecules, and their transport is energetically coupled with an electrochemical gradient (typically H^+ or Na^+) (Höglund *et al.,* 2010; Yan, 2015). Pit proteins are low-affinity systems with a K_M that varies greatly (1-1000μM Pi) in comparison to the Pst system (Bøttger and Pedersen, 2011)

1.2 - Pit Inorganic Phosphate Transport Systems

Pit systems have been studied primarily in *E. coli*, but also recently in eukaryotes like *Homo sapiens* and *Saccharomyces cerevisiae*. Pit proteins are identified by the PHO4

domain (Pfam: PF01384), which refers to the PHO-4 Na(+)/Pi symporter from *Neurospora crassa* (NcPHO-4) (Versaw and Metzenberg, 1995). Pit proteins typically contain 10-12 transmembrane helices with two signature PHO4 sequences that are duplicated between the two-halves of the protein. This structure is thought to have arisen via a tandem duplication event hypothesized to have occurred early in the evolutionary history of MFS proteins (Saier *et al.*, 2000; Bøttger and Pedersen, 2011; Reddy *et al.*, 2012) (Figure 1-3).

1.2.1 - *E. coli* PitA system

The canonical bacterial Pit protein is *E. coli* PitA (EcPitA) which has been studied since the early 1970s using an *E. coli pst* mutant (Medvecsky and Rosenberg, 1971). EcPitA is a protein of 499 a.a. with a K_M of 30-40 µM Pi (Rosenberg *et al.*, 1977; Elvin *et al.*, 1986). The transport reaction of EcPitA is energized by proton-motive force which is required for uphill transport of Pi with up to three protons, and possibly includes a cation (K^+, Ca^{2+} or Mg^{2+}), which together create a neutral-metal substrate ($HMePO_4$) (Rosenberg *et al.*, 1979; Russell and Rosenberg, 1979; Russell and Rosenberg, 1980; van Veen *et al.*, 1944). Unlike the Pst system, EcPitA is also observed to transport arsenate, due to the similarities between Pi and arsenate (AsO_4^{-3}) (Willsky and Malamy, 1980). Additionally, exchange of Pi is observed to occur via EcPitA, as the addition of unlabeled substrate (either Pi or AsO_4^{-3}) resulted in the exit of Pi from the cells, however the downhill exit of Pi is not observed (Willsky and Malamy, 1980). The expression of

EcPitA has been proposed to be induced by Zn^{2+} or other ions but this remains inconclusive (Beard *et al.*, 2000).

1.2.2 - Eukaryotic Pit systems

The human Pit1 and Pit2 (HsPit1 and HsPit2, respectively) orthologues have also been studied, particularly due to their association with mineralization diseases of the brain and renal systems as well as osteoblast-associated diseases (Jono *et al*., 2000; Wang *et al.*, 2012). Pit1 and Pit2 are both low-affinity Pi transporters (K_M of 24-40 µM Pi) that are coupled with the transport of Na^+ with a Pi molecule (Salaün *et al.*, 2001; Bøttger and Pedersen, 2011). Interestingly, HsPit proteins are also demonstrated to be receptors of the murine leukemia virus, which was used to predict the membrane topology and demonstrate membrane localization of the proteins (Kavanaugh *et al*., 1994; Salaün *et al.*, 2001). Additionally, HsPit2 has been shown to oligomerize *in vivo*, using antibodies generated against the intracellular (cytoplasmic) regions of the protein (Salaün *et al.*, 2002).

Similarly, *S. cerevisiae* has a diverse set of high- and low-affinity Pi transport proteins that have optimum activity in a wide range of environmental conditions (Samyn *et al.*, 2016). *S. cerevisiae* Pho89 (ScPho89) is a major Pi uptake system that is a member of the Pit family with homology to the Na(+)-Pi symporter HsPit1/HsPit2 (Zvyagilskaya *et al*., 2008). Pho89 displays optimum Pi uptake at alkaline pH and with Na(+) ions, similar to that of NcPHO-4 (Sengottaiyan *et al*., 2013).

Mutational analysis of the signature PHO4 motifs in HsPit2 as well as ScPho89 reveal that highly-conserved glutamic acid (G) residues are essential for Pi uptake via these system (Bøttger and Pedersen, 2002; Bøttger and Pedersen, 2011; Samyn *et al.*, 2012). Similarly positioned glutamic acid residues are also present in prokaryotic Pit proteins, but their relevance and function in Pi uptake has not been directly explored.

1.2.3 - *Sinorhizobium meliloti* Pap-Pit system

Bacteria with mutations in the *phnCDET* genes of *S. meliloti* strain 1021 are defective in symbiotic N_2-fixation, and second-site mutations that suppressed the Fix^- phenotype of *phnCDET* mutants were mapped to a locus that was subsequently found to encode the *pap-pit* genes (Oresnik *et al*., 1994, Bardin *et al.*, 1998). The locus was originally designated as *orfA-pit* and because of the close association between *orfA* and *pit* genes in many organisms (see below), we re-annotated *orfA* as *pap* for Pit associate protein. Three suppressor mutations were identified as single thymidine deletions within a hexa-thymidine sequence in the *pap-pit* promoter region which increased transcription of the *pap-pit* genes. The *pap-pit* genes were found to encode a Pi uptake system with a K_M of 1-2µM Pi (Voegele *et al.*, 1997). The *pap* and *pit* genes appear to form an operon with their coding sequences overlapping by a single base pair. Transcription of *pap-pit* is repressed under low Pi growth conditions in a PhoB-dependent manner.

S. meliloti pit encodes a protein of 334 a.a., which is almost 200 a.a. smaller than EcPitA (Figure 1-2). Several *pit* orthologues from *Shewanella oneidensis* and *Bacteroides*

thetaiotaomicron, are also located downstream of a *pap*-like gene, and these Pit proteins are also smaller than EcPitA.

1.3 - Pap and PhoU

The structures of Pap proteins from *Bacteroides thetaiotaomicron* and *Shewanella oneidensis* (BtPap and SoPap, respectively) have been determined by X-ray diffraction and deposited in Protein Data Bank (Joint Center for Structural Genomics; PDB: 3L39 and 2OLT, respectively). While no dedicated publications are associated with these structures, both structures are intriguingly similar to the crystal structures of PhoU from other organisms (Liu *et al.,* 2005; Oganesyan *et al.*, 2005). Both PhoU and Pap forms bundles of 6 parallel α-helices separated by short loops and both are predicted to form dimeric or trimeric complexes. Additionally, PhoU proteins have two PhoU domains (Pfam: PF01895) with proposed metal-binding motifs ([E/D]XXXD motifs) that interact with Fe clusters (Liu *et al.*, 2005).

PhoU has been studied due its connection to the Pho regulon and PstSCAB transport (Steed and Wanner, 1993). PhoU is believed to function as a negative regulator of the Pho regulon via interactions with PhoR-PhoB. Mutations in *phoU* cause constitutive Pho regulon expression (Surin *et al*., 1986).

PhoU may function as a modulator of Pi uptake via the PstSCAB system (Rice *et al.,* 2009; diCenzo *et al*., 2017). *phoU* mutants accumulate polyphosphate (polyP), and grow poorly and second-site suppressor mutations are commonly found within the *pst* genes (Morohoshi *et al.,* 2002; de Almeida *et al.,* 2015; Lubin *et al.,* 2015; diCenzo *et al*.,

2017). Additionally, there is some evidence that the modulation of PstSCAB by PhoU occurs through a direction interaction between PhoU and PstB, while also forming a signaling complex with PhoR-PhoB, which presents a bi-functional model for modulation of Pi uptake and Pho Regulon (Gardner *et al.*, 2014).

Although there is a multitude of experimental evidence indicating PhoU plays a critical role in Pi metabolism (via the Pho regulon or PstSCAB), the exact mechanism through which PhoU acts is not known. The presence of the metal-binding motifs may suggest that PhoU has some catalytic activity.

1.5 – Research Objective

The overall objective of this work was to determine the function of the Pap protein in Pi-uptake via the Pap-Pit Pi-transport system. The function of the Pap protein was explored in relation to its cognate Pit transporter. This was conducted through a deletion-complementation approach, observing phenotypic effects upon growth (in minimal media with Pi as the sole source of P) and radiolabeled-Pi uptake experiments.

Pap is hypothesized to be a modulator of Pi uptake via Pit as, (i) *pap* and *pit* genes commonly occur together, similar to the *pstSCAB* operon and *phoU* genes, and
(ii) *pap* mutations result in defective Pi uptake, in contrast with *phoU* mutations that cause excessive Pi uptake via PstSCAB.

Because of the structural similarities between Pap and PhoU and their association with Pi transport, a common mechanism-of-action for these proteins is proposed. Hence the function of the Pap protein may provide intriguing information regarding the function of PhoU and PhoU-like proteins.

Chapter 2 – Materials and Methods

2.1 - Growth media, culture growing conditions and antibiotics

The bacterial strains and plasmids utilized in this study are listed in Table 2-1. Where mentioned, isopropyl β-D-1-thiogalactopyranoside (IPTG) was included at a final concentration of 0.5mM for broth media and 1mM for agar. Glycerol 3-phosphate (G3P) was added to a final concentration of 1mM.

2.1.1 - Luria-Bertani (LB) media

LB was prepared with 1% (w/v) Bacto™ Tryptone, 0.5% (w/v) Bacto™ Yeast Extract, and 0.5% (w/v) NaCl mixed with MilliQ water and sterilized by autoclaving. LB agar was prepared with 1.5% (w/v) Difco agar. LBmc was LB supplemented with 2.5mM $MgSO_4$ and $CaCl_2$.

2.1.2 - MOPS-buffered and M9 minimal media

MOPS-buffered minimal medium was composed of 40 mM MOPS/20mM KOH, 8.6 mM NaCl, 18.7 mM NH_4Cl, 1 mM $MgSO_4$, and 0.25 mM $CaCl_2$. Additionally, the media was supplied with 1 µg/mL thiamine-HCl, 0.3 µg/ml biotin, 10 ng/ml $CoCl_2$, 38 µM $FeCl_2$ with the addition of 20 mM glucose as carbon source. K_2HPO_4 was added to final concentrations of 2mM (MOPS-P2). M9 minimal medium was composed of Difco™ M9 Minimal Salts (final concentration of 3.39% (w/v) Na_2HPO_4, 1.5% (w/v) KH_2PO_4, 0.25% (w/v) NaCl, 0.5% NH_4Cl) supplemented with 1 mM $MgSO_4$, 0.25 mM

$CaCl_2$, 10 ng/mL $CoCl_2$, and 1 µg/mL thiamine-HCl. Mannitol or glucose (20 mM) was added as a carbon source.

2.1.3 - Antibiotics

Where necessary, for *S. meliloti*, solid media were supplemented with appropriate antibiotics at following concentrations: streptomycin (Sm), 200 µg/ml; gentamycin (Gm) 60 µg/ml; neomycin (Nm) 200 µg/ml; spectinomycin (Sp) 200 µg/ml. *E. coli* was used for routine cloning and was grown at 37°C in LB medium with antibiotics at the following concentrations: gentamycin (Gm) 10 µg/ml; kanamycin (Km) 20 µg/ml; spectinomycin (Sp) 100 µg/ml; chloramphenicol (Cm) 20 µg/ml. For all strains, antibiotic concentrations were reduced to half while growing in liquid media.

2.2 - Molecular cloning and recombinant DNA techniques

2.2.1 - Polymerase Chain Reaction (PCR)

A comprehensive list of primers is included in Appendix Table 1. All oligonucleotides were synthesized by Integrated DNA Technologies Inc. (Coralville, Iowa, USA). DNA pellets were resuspended and dissolved with ddH_2O to a final working concentration of 10µM. PCR amplification was conducted using *i-pfu* DNA polymerase (iNtRON Biotechnology, Seongnam, South Korea) as per the manufacturer recommendations, unless otherwise stated. Forward and reverse primers were used at a final concentration of 0.3µM. dNTPs were used at a final concentration of 0.2µM.

Genomic DNA template was used at a concentration of 20ng/µL and plasmid DNA template was used at 5-10ng/µL. Melting temperatures (T_m) was determined using the UC Berkeley Oligo Calculator (http://mcb.berkeley.edu/labs/krantz/tools/oligocalc.html). PCR products were subsequently purified using a DNA Purification Kit (GeneAid #DFH300, Taipai, Taiwan) prior to further use.

2.2.2 - Plasmid DNA preparation, digestion and ligation

Plasmid DNA was prepared from *E. coli* DH5α using the Plasmid Mini-Prep Kit (GeneAid, #PDH300, Taipai, Taiwan) to the manufacturer recommendations. A single aliquot of elution buffer (10mM Tris-HCl pH 8.0) was passed through the column several times to concentrate the plasmid DNA. Concentrations were determined by UV spectrophotometry at absorbances of 260/280nm (Biotek Cytation 3 96-well plate Spectrophotometer).

DNA digestion was conducted using New England Biolabs (NEB)-supplied restriction enzymes to manufacturer recommendations. DNA was no more than 25% of the total reaction volume. Digestion reactions were subsequently purified using a DNA Purification Kit (DFH300) prior to further use.

Conventional ligation reactions were conducted in a 10 µL reaction (with insert:vector molar ratio of 1:4 for two fragments, or 1:2:2 for two fragments) using T4 DNA ligase (New England Biolabs). Reactions were incubated at room temperature (~20^{o}C) for 30 minutes to permit ligation of the fragments. Alternatively, sequence-ligation independent cloning (SLiC) was also used for cloning, where mentioned (Jeong

et al., 2012). Then, DNA was added to chemically competent *E. coli* DH5α, and following 30 minutes on ice, cells were heat shocked at 42^{o}Cfor 45 seconds, followed by 2 minutes on ice. 1 mL LB broth was added and after a 1-hour incubation at 37^{o}C, transformants were then spread onto selective LB agar plates.

Putative transformants were inoculated in LB broth with appropriate antibiotics and incubated overnight and the plasmid DNA was purified the following day. Putative clones were screened by a combination of restriction digestion mapping (using restriction sites within and outside the insert) to confirm the presence of the insert DNA. Alternatively, colony PCR using flanking primers annealing to the vector were used to confirm the presence of insert DNA in the plasmid. Furthermore, all constructs were sequence verified by Sanger sequencing at McMaster MOBIX Sequencing Facility (McMaster University, Hamilton, Canada).

2.2.3 - Annealing and ligation of synthesized oligonucleotides

For the addition of small peptide tags, equimolar amounts of two complementary oligonucleotides were annealed in a reaction consisting of 100 mM Tris-HCl (pH 8.0), 1 mM EDTA and 62.5 mM NaCl. Negative control reactions were also conducted which contained only one of the two oligonucleotides. The reaction was then incubated at 95 ^{o}C for 2 minutes before ramping down 1 ^{o}C per minute until the expected T_m of the oligos. This T_m value was then held for 2 min before ramping down 1 ^{o}C per minute until 4 ^{o}C. The newly annealed dsDNA fragment was then confirmed by visualization on a 2% agarose gel at run 100 V for 30 minutes.

2.2.4 - Overlap Extension PCR

Where mentioned, multi-insert cloning was conducted by overlap extension PCR. Primers were generated to carry 5' overhangs with at least 30bp homology between the insert fragments to be spliced together. First, PCR products were amplified using the sequence of the template-annealing portion of the primer to calculate the annealing temperature. Next, the fragments were joined via overlap reaction: equimolar amounts of insert DNA (0.1-1pmol) and a final concentration of 0.2µM dNTPs were mixed in a PCR tube. The reaction was then run through 15 cycles of denaturation, annealing and extension, where the annealing temperature was the annealing homology regions between the fragments and the extension time was based on the size of the largest fragment. Afterwards, an overlap PCR using 5µL of overlap reaction as the template DNA in a 50µL reaction with the end primers of the total joined product. As a negative control, the overlap extension and overlap PCR using only one of each insert DNA fragments were used to determine whether any self-annealing occurred which would rule out any non-specific products. All reactions were then run on an agarose gel and the expected band from the overlap extension PCR was cut out and gel purified.

To generate point mutations, a similar method was utilized where two insert DNA fragments were amplified with 30bp homology using the 5' overhangs of the primers. The 5' overhangs included complementary mutations that would result in a specific amino acid change in the result product. The fragments were joined by overlap extension PCR

and the subsequent clones were sequence verified to confirm that the expected mutation and no other mutations were present.

2.4 - General transduction, plasmid conjugal mating, and homologous recombination in *S. meliloti*

General transduction of *S. meliloti* and conjugation of plasmids from *E. coli* DH5α into *S. meliloti* were conducted as previously (Milunovic et al., 2014).

2.5 - Previous work on *S. meliloti* Pap-Pit

Previous work on the *S. meliloti* Pap-Pit project was conducted by Dr. Rahat Zaheer, Hui Zhao and Dr. Vahid Husseininaveh.

2.5.1 - Constructing an *S. meliloti* Δ*pap* and Δ*pit* mutant strain

To construct a *pap* (*smc02862*) deletion mutant, two fragments containing regions upstream (including the first 18bp) and downstream (including the last 27bp) of the *pap* coding sequence were amplified (using primers orfAkn-F1/orfAkn-R1, and orfAkn-F2/orfAkn-R2) were cloned into pJQ200 (Quandt & Hynes, 1993) and subsequently transformed into *E. coli* DH5α to create pTH2340. The expected deletion of the *pap* coding sequence resulted in a Pap protein deletion that retained only the first 6 and last 8 amino acids. A similar strategy using primers pitkn-F1/pitkn-R1 and pitkn-F2/pitkn-R2 was employed to construct a plasmid (pTH2341) for deleting the *S. meliloti pit* gene (*smc02861*) (Figure 2-2).

Positive plasmid clones were mated into *S. meliloti* RmP110 strain and single cross-over transconjugants were selected on LBmc(SmGm) media. Representative transconjugants were inoculated in 5mL of LBmc and grown overnight at 30^{o}C. Cultures were serially diluted and dilutions were plated onto LB plates containing 5% sucrose to select for recombinants that lost the vector containing the *sacB* gene. One hundred colonies from the sucrose plates were patched onto LB(Sm) with 5% sucrose and subsequently on LB(SmGm) to identify Sm-resistant Gm-sensitive recombinant colonies.

RmP1628 (RmP110 Δ*pap*) and RmP1629 (RmP110 Δ*pit*) deletion mutants were then verified by colony PCR and sequencing.

2.5.2 - Constructing an *S. meliloti* Δ*pap-pit* mutant strain

The deletion of *pap-pit* was generated by amplifying *the pap-pit* locus with 500bp upstream and 500bp downstream of the coding sequences (using primers 500-PapF and PitR-500) and inserting this fragment into pUCP30T via the XbaI/HindIII restriction sites to create pTH2824. This plasmid was then transformed into *E. coli* BW25113 carrying pKD46 (that expressed λ-red recombinase under L-arabinose induction) to create M2095. Then, an FRT-Km-FRT cassette was amplified with 30bp homology to the immediate upstream and downstream regions of the *pap-pit* coding sequences. This PCR product was electroporated into electrocompetent M2095 (grown with L-arabinose) followed by plating on LB(GmKm) agar, selecting for transformants with a double-recombination to create pTH2897, which had the *pap-pit* coding sequences replaced by a Km cassette. This vector was then mated into RmP110, selecting for double recombinants on LBmc(SmNm) medium. To eliminate the Km/Nm cassette, the *flp* recombinase plasmid pTH1944 (Milunovic et al., 2014) was mated into the previous strain and screened for sensitivity to Nm.

2.5.3 - Constructing *S. meliloti pap* point mutations

pap point mutations were constructed in pUCP30T by overlap extension PCR of two fragments (Figure 2-3). The forward primer of the first fragment (500Pap-F) and

reverse primer of the second fragment (PitR-500) were common primers used for the construction of all point mutations, with internal primers which were complementary and shared mutations resulting in an alanine transition mutation. Once the two fragments were PCR amplified independently, they were spliced together by overlap extension PCR using 500Pap-F and PitR-500 as the end primers. The final PCR product was cloned into pUCP30T via XbaI/HindIII restriction sites. Table 2-4 lists plasmids containing *pap* point mutations.

2.6 - Conditionally phosphate uptake-deficient mutant of *S. meliloti*

An *S. meliloti* $lacI_q$, P_{lac}::*pstSCAB-phoUB* and P_{lac}::*phoR* and *phnC*::Sp^R strain (RmP3421) was generated and was the background strain used for studying the effects of *pap* and/or *pit* mutations on Pit-dependent phosphate uptake (diCenzo *et al.,* 2017). IPTG was included during routine culturing of this strain and derivative strains for induction of the Pho regulon and PstSCAB system in order to avoid second site suppressor mutations that would otherwise restore the Pi uptake or the Pho regulon in a *pit, pst* and *phn* triple mutant.

2.6.1 - Transducing *pap-pit* deletions

A *S. meliloti* phenylalanine auxotroph (RmG430 *phe-232::*Gm^R), likely *pheA* (*pheA* lies 40 kb downstream of *pap-pit*) was used as a recipient to transduce the unmarked *pap-pit* deletions into RmP3421 (wildtype *pap-pit*) (Figure 2-5). First, the *phe-232::*Gm^R allele was transduced from RmG430 into RmP3421 to create RmP3923 by

selecting transductants on M9-glucose containing Gm and 1mM phenylalanine. RmP3923 transduced with φM12 lysates from RmP1628, RmP1629, and RmP3151, and transductants were selected on M9-glucose without phenylalanine and subsequently screened for sensitivity to gentamycin. These transductants produced strains RmP3924 (RmP3421 Δ*pap*) RmP3925 (RmP3421 Δ*pit*) and RmP3926 (RmP3421 Δ*pap-pit*), respectively, which were all confirmed by colony PCR and sequencing of the deletion region.

2.6.2 - Complementation of *pap-pit* deletions via homologous recombination

To complement these deletion mutants, regions of *the pap-pit* locus were amplified and cloned into the suicide vector pUCP30T via XbaI/HindIII, to generate single-crossover recombinants into the *S. meliloti pap-pit* deletion loci (Figure 2-6). For the *pap-pit*^{truncated} construct (promoter region with *pap* coding sequence and the first 300 bp of the *pit* coding sequence), recombination was expected to occur with the Δ*pap* loci, where a single-crossover at the truncated *pit* homologous region would insert the wildtype *pap* coding sequence upstream of the chromosomal *pit* gene. And similarly, recombination of the *pap*^{truncated}*-pit* construct (last 318 bp of the *pap* coding sequence and the entire *pit* coding sequence) with Δ*pit* was expected to occur to insert the wildtype *pit* coding sequence downstream of the chromosomal *pap* gene. All complementation constructs were then mated into RmP3421, RmP3924, RmP3925 and RmP3926 to determine whether the integration complemented the mutant phenotypes (Figure 2-7).

2.6.3 - Construction of recombinant *S. meliloti* Pit~tag fusion

Membrane topology of integral membrane proteins can be determined by the fusion of reporter proteins to truncated proteins, where activity of the reporter protein would be indicative of the intracellular or extracellular localization of the putative loop regions. Alternatively, the reporter activity can be used to detect the presence or absence of the recombinant protein. pPLE01 (pBluescript II SK (+) (Amp^R) carries a recombinant alkaline phosphatase-β-galactosidase (*phoA/lacZ*) fusion, which also has distinct restriction enzyme sites upstream and downstream of the fusion (Islam *et al*., 2010). A recombinant *S. meliloti* Pit~PhoA/LacZ fusion was constructed, where alkaline phosphatase activity was used as a reporter of the presence of Pit.

The wildtype *pap-pit* operon was cloned within pPLE01, upstream of the *phoA/lacZ* fusion via PstI, creating a C-terminal translational fusion of Pit~PhoA/LacZ with a 6 a.a. linker region. Next, 500bp downstream of the *pit* gene was PCR amplified and cloned immediately downstream of the *phoA/lacZ* gene via HindIII.

The resulting *pap-pit~phoA/lacZ* fusion was subcloned into pTH2824 (wildtype *pap-pit*) via StuI and BspEI restriction sites, selecting for transformants on LB Gm to create pTH3198. Next, the promoter-*pit* operon was subcloned from pTH3173 into pTH3198 via HindIII and StuI restriction sites to create pTH3196, to create a recombinant fusion for Pit~PhoA/LacZ lacking the *pap* gene.

Small peptide tags were also used for the detection of *S. meliloti* Pit protein via Western Blot. A variety of peptide tags were generated to the C-terminus of the Pit protein by a combination of techniques (Figure 2-8). The complementary

oligonucleotides were annealed to generate short DNA fragments (~70bp) that contained the amino acid sequences of the small peptide tag: bovine rhodopsin 1D4: TETSQVAPA (Oligos: 1D4-F, 1D4-R), FLAG: DYKDDDDK (Oligos: FLAG-F, FLAG-R), Strep-II: WSHPQFEK (Oligos: StrepII-F, StrepII-R). An in-frame linker amino acid sequence of the TEV cysteine protease motif (ENLYFQ) was also included so that the tag could be later removed if necessary. Annealing of the oligonucleotides created 4 bp overhangs on both ends that match the digested PstI restriction product.

The short fragments were then cloned into pUCP30T (*pap-pit~phoA/lacZ*) which has two PstI restriction sites, immediately downstream of the *pit* coding sequence and after the *phoA/lacZ* reporter protein, selecting on LB Gm and screening for loss of alkaline phosphatase activity. Sequencing was used to confirm the orientation of the tag was appropriate to create a C-terminal translational fusion. The resulting constructs were *pap-pit*~Rho1D4 (pTH3200), *pap-pit*~FLAG (pTH3201), *pap-pit*~StrepII (pTH3202) constructs (Figure 2-8). These constructs were subsequently mated into RmP3926, selecting for transconjugants on LBmc (SmGm) IPTG agar.

Alternatively, a fragment of *pit* gene was amplified with a reverse primer that contained a 5' overhang region with a poly-histidine (His10) peptide tag (Oligos: AgeI-*pit*F, pitHis10R) and the 500bp region downstream was amplified with a forward primer that contained the complementary 5' overhang to the His10 overhang (Oligos: pitHis10F, pitR-HindIII) (Figure 2-9). These two fragments were amplified and joined by overlap extension PCR before being cloned into pTH2824 via AgeI and HindIII restriction sites.

2.6.4 - Phenotype of Pap missense mutations

To determine the effect of *pap* point mutations on phosphate uptake via Pit, the pUCP30T vectors which carry *pap-pit* with point mutations were mated into RmP3926 (Δ*pap-pit*) selecting for single recombinants on LBmc (SmGm) with 1 mM IPTG. Single colonies were then purified by streak twice onto the same media before growth and phosphate uptake experiments

2.7 - Pi-uptake deficient *E. coli* MT2006

A phosphate uptake-deficient strain of *E. coli* was created by Motomura *et al.* (2011) which carried deletions of all phosphate uptake systems via *flp* recombinase-FRT of a kanamycin resistance cassette to create MT2006 (MG1655 *ΔpitA::frt ΔpitB::frt ΔphnC::frt Δ(pstSCAB-phoU)560::Km*^{R}). This strain is unable to grow with inorganic phosphate as the sole source of phosphorus, which indicates there are no functional phosphate uptake systems. Thus, a plasmid which complements this mutant phenotype carry genes encoding functional phosphate uptake systems.

For routine experimentation, 1 mM glycerol 3-phosphate (G3P) was included as an alternative source of phosphorus. G3P is transported into the cell via GlpT or Ugp systems, which are both present in the cell.

2.7.1 - Subcloning *pap* alleles into pUC119

The *pap* point mutations in pUCP30T were subcloned into the high copy-number vector, pUC119, via MluI/AgeI cut sites. The plasmid DNA was sequenced to confirm

that subcloning by digestion did not result in any unexpected mutations within the *pap* gene. These plasmids were then transformed into chemically competent *E. coli* MT2006, selecting for transformants on LB Amp G3P.

2.7.2 - Cloning of heterologous *pap-pit* systems in pMW119

pap-pit orthologues from *Sinorhizobium meliloti* Rm1021 (SmPapPit, SMc02862 and SMc02861), *Bacteroides thetaiotaomicron* VPI-5482 (BtPapPit, BT_4638 and BT_4637), and *Shewanella oneidensis* MR-1 (SoPapPit, SO3770 and SO_3771) were cloned into the low copy number vector pMW119 via HindIII/XmaI restriction sites under control of the *lac* promoter (Figure 2-10). *E. coli* MG1655 PitA (EcPitA, b3493) was also cloned into pMW119 as a positive control for growth and Pi uptake complementation, as well as a comparison between Pap-Pit and PitA systems. PCR amplified fragments included 15-25 bp upstream and downstream of the native gene(s). Additionally, the *pit* genes (including 15-25bp upstream of the *pit* start codon) were amplified and cloned into pMW119 via the HindIII/XmaI restriction sites. These vectors were subsequently transformed into MT2006, by selecting on LB Amp with 1mM glycerol 3-phosphate (G3P).

2.8 - Growth complementation experiments

Growth experiments were conducted to determine whether the phosphate uptake-deficient mutant phenotypes of *E. coli* or *S. meliloti* could be complemented by different constructs.

The initial preparation of *S. meliloti* and *E. coli* varied and is described below (2.8.1 and 2.8.2). To begin the growth experiments, 15µL of cells were aliquoted in technical triplicates onto a 96-well plate (Corning In., Durham, NC, USA), followed by 135µL of the appropriate media, such that the starting OD_{600nm} would be ~0.05. Growth was conducted using a BioTek^{Tm} Cytation 3 plate reader with double-orbital shaking (1mm radius, 807 cycles per min) and temperature set at 30^{o}C for *S. meliloti* strains, whereas *E. coli* experiments were conducted at 37^{o}C. The OD_{600nm} of each well was read every 15 minutes and the average of the three triplicates was plotted and analyzed using Microsoft Excel.

2.8.1 - Conditionally phosphate uptake-deficient *S. meliloti*

Single colonies of *S. meliloti* were inoculated in LBmc with 0.5mM IPTG and incubated overnight at 30^{o}C. Cells were then harvested by centrifugation, washed three times and resuspended with 1mL MOPS-P2 (0mM IPTG). The cells were then depleted of PstSCAB-PhoUB and PhoR proteins by subculturing into 5mL MOPS-P2 (0mM IPTG) to an OD_{600nm} of ~0.1 and incubated at 30^{o}C for 16 hours. Prior to the growth measurements, cells were harvested by centrifugation and washed once and resuspended with 1mL MOPS-P2 (0mM IPTG) to an OD_{600nm} of ~0.5.

2.8.2 - Pi-uptake deficient *E. coli*

Single colonies were inoculated in LB with 1mM G3P and incubated overnight at 30^{o}. Cells were then harvested by centrifugation then washed three times and resuspended

with M9 without G3P. The cells were then depleted of intracellular Pi by subculturing into 5mL M9 20mM mannitol to an OD_{600nm} of ~0.4 and incubated at 37°C for 2 hours. Prior to the growth measurements, the cells were harvested by centrifugation, washed once and resuspended with M9 media to an OD_{600nm} of ~0.5.

2.9 - Polyacrylamide gel electrophoresis and Western Blot

Polyacrylamide gel electrophoresis (PAGE) and western blot (WB) were used to detect the *S. meliloti* Pap and Pit proteins.

Denaturing SDS polyacrylamide gels were composed of a 10 or 12 % resolving gel (Tris HCl pH 8.8) and a 4 % stacking gel (Tris-HCl pH 6.8) using 30 % Acrylamide/Bis solution, 37.5:1 from Bio-Rad Laboratories, Inc. (#1610158). Gels were cast and electrophoresis was run using the Bio-Rad Mini-PROTEAN systems at 200 V for 30-40 minutes or until the dye reached the bottom of the gel. Protein samples were mixed with 2X Laemmli buffer (65.8 mM Tris-HCl pH 6.8, 30% (v/v) glycerol, 4% (w/v) SDS, 100 mM DTT, 355 mM 2-mercaptoethanol, 0.02% (w/v) bromophenol blue).

2.9.1 - Protein isolation from *S. meliloti*

Single colonies were inoculated in LB with appropriate antibiotics and grown to OD_{600nm} of ~1-2. The cells were harvested and washed twice with phosphate-buffered saline (PBS; 10mM NaH_2PO_4, 1.8mM KH_2PO_4, 137mM NaCl, 2.7mM KCl, 0.15mM phenylmethylsulfonyl fluoride (PMSF) and 50mM dithiothreitol (DTT)). Alternatively,

the cells from a completed growth experiment were harvested and washed twice with PBS.

Alternatively, cell suspensions of equivalent OD_{600nm} were sonicated using a microtip sonicator: 10 seconds on (40 % duty cycle), 3 minutes on ice, for 5-8 cycles or until the cell suspension became translucent. The lysate was then centrifuged at 5,000 g for 10 minutes at 4 ^{o}C to precipitate unlysed cells and cell debris. The protein concentration of cleared lysates was determined by the Bradford Method prior to running equivalent amounts of protein on SDS-PAGE and detection via WB.

2.9.2 - Sub-cellular fractionation

To determine the sub-cellular localization of Pap and Pit proteins, the insoluble and soluble protein fractions were separated by centrifugation. Cell lysates were centrifuged at 21,000 g for 2 hours at 4^{o}C to precipitate the insoluble fraction, which appeared as a red-brown translucent pellet. The supernatant (containing the soluble fraction) was separated into a new tube. The insoluble fraction was then washed twice with PBS without disrupting the pellet. Then, the pellet was resuspended using PBS with either 0.5% (w/v) SDS, 0.4% (v/v) NP-40, or 2% (v/v) Triton X-100. The pellet was slowly dissolved at room temperature with occasional pipetting for 20-30 minutes or until the pellet was no longer visible.

2.9.3 - SDS-PAGE

Cell suspensions of equivalent OD_{600nm}, protein lysates or cytosolic protein fractions were mixed with 2X Laemmli sample buffer (65.8 mM Tris-HCl, pH 6.8; 20% (w/v) glycerol; 4% SDS; 0.01 % bromophenol blue) and boiled at 95 ^{o}C for 10 minutes (. The boiled lysate was then centrifuged at 5,000 g for 10 minutes at 4 ^{o}C to precipitate unlysed cells and cell debris. The supernatant was loaded onto the gel and run at 200 V for 30-40 minutes or until the dye reached the bottom of the gel.

2.9.4 - Western Blot

Proteins from the SDS-PAGE were transferred onto a Immun-blot polyvinylidene difluoride (PVDF) membrane (Bio-Rad Laboratories, Ltd.) with the Trans-blot SD Semi-Dry Transfer Cell (Bio-Rad Laboratories, Ltd.) The gel was soaked in transfer buffer (4.8mM Tris, 0.39mM glycine, 20% (v/v) methanol) for at least 30 minutes prior to transfer. The PVDF membrane was activated with 80% methanol for 1 minute before equilibration in the transfer buffer for at least 5 minutes. Two 5-millimeter filter papers were also soaked in transfer buffer for at least 5 minutes. The filter-membrane-gel-filter 'sandwich' was then placed on the transfer apparatus and proteins were transferred at 10V for 90 minutes.

After blotting, the membrane was soaked in 5% skim milk blocking buffer overnight at 4^{o}C. The following day, the membrane was washed three times with Tris-buffered saline with Tween-20 (TBST; 10mM Tris-HCl pH 8.0, 15mM NaCl, 0.5% Tween-20) for 5 minutes with shaking. The membrane was then incubated with rabbit

anti-Pap antisera (1:5000 dilution) for 2 hours with shaking at room temperature. The membrane was again washed three times with TBST before incubation with goat anti-rabbit alkaline phosphatase conjugate antibody (1:3000 dilution) for 2 hours with shaking at room temperature. The membrane was washed three times with TBST before equilibration with alkaline phosphatase (AP) buffer (10mM Tris-HCl pH 9.0, 10mM NaCl, 0.5mM $MgCl_2$) for 5 minutes. The blot was then developed using alkaline phosphatase buffer with 0.33mg/mL nitro blue tetrazolium chloride (NBT) and 0.17mg/mL 5-bromo-4-chloro-3indolyl phosphate (BCIP) in the dark for 30 to 60 minutes.

2.9.5 - Protocol alterations for detection of recombinant Pit protein

For detection of recombinant *S. meliloti* Pit, the protein samples were mixed with 2 X Laemmli buffer and incubated at 30 ^{o}C for 30 minutes with slight mixing. To prevent formation of protein aggregates, samples were not boiled prior to separation by SDS-PAGE at 100 V for 60-70 minutes (Privé, 2007).

For semi-dry blotting of recombinant Pit onto the PVDF membrane, the membrane and gel were incubated in modified transfer buffer (15.6 mM Tris, 120 mM glycine, 10 % (v/v) methanol and 0.02 % (w/v) SDS) to improve transfer of hydrophobic proteins. The transfer was conducted at 10 V for 2-3 hours at 4 ^{o}C to prevent overheating of the membrane. Primary and secondary antibodies were incubated at 4 ^{o}C overnight for each step rather than at room temperature. Development of the blot image was conducted with the same solution but for several hours.

2.9.6 - *in vivo* protein-protein cross-linking using formaldehyde

Interactions between *S. meliloti* Pap and other proteins were screened by *in vivo* protein cross-linking using formaldehyde. Cells were inoculated in LB with appropriate antibiotics and grown to the required density. The cells were then washed twice with PBS. An equivalent of 1mL of cells at OD_{600nm} ~1.0 was aliquoted and pelleted by centrifugation. The pellet was then resuspended with PBS containing 1% (v/v) formaldehyde and incubated at room temperature with gentle agitation for 8 minutes. The cells were then centrifuged for 2 min at 5,000 g. The supernatant was discarded and the pellet was resuspended in 1mL ice-cold PBS with 1.25M glycine and incubated on ice for at least 5 minutes. The pellet was then washed again with PBS with 1.25M glycine before being resuspended in PBS and lysed by boiling with 2X Laemmli buffer.

2.9.7 - Native-PAGE and in-gel zymogram of Pit~PhoA/LacZ

A non-denaturing continuous 10 % polyacrylamide gel (TBE pH 8.7) was used to resolve proteins in native conformation. Protein samples were mixed with 3 X Native PAGE sample buffer (5mM Tris-2.5mM boric acid, pH 8.7; 30% glycerol, 0.03% bromophenol blue). The non-denaturing PAGE was then run at 150V for 90 minutes or until the dye reached the bottom of the gel.

To detect alkaline phosphatase activity of the Pit~PhoA/LacZ fusion, the non-denaturing gel was by equilibrating in AP buffer (10mM Tris-HCl pH 9.0; 10mM NaCl, 0.5mM $MgSO_4$) for 5-10 minutes before incubating the gel with AP buffer with NBT and

BCIP (Final concentrations of 5mg/mL 5-bromo-4-chloro-3-indolyl phosphate, BCIP; 10mg/mL nitro blue tetrazolium chloride, NBT) for 20-30 minutes in the dark.

2.10 - Phosphate uptake experiments

Cells were inoculated in liquid medium that permitted phosphate uptake or assimilation: LBmc 0.5 mM IPTG for *S. meliloti* recombinants and LB 1 mM G3P or LB 1 mM G3P 0.5 mM IPTG for *E. coli* MT2006 transformants. Inocula were incubated at 30 ^{o}C overnight (16-24 hours) with shaking. The following day, cells were harvested, washed three times with 1 mL MOPS-P0. For *S. meliloti,* the cells were subcultured into 15 mL of MOPS-P2 to a final OD_{600nm} of ~0.1 and incubated at 30 ^{o}C for 16 hours with shaking to deplete the cells of the PstSCAB-PhoU proteins. For *E. coli,* the cells were subcultured into 15 mL of MOPS-P0 to a final OD_{600nm} of ~0.1 and incubated at 37 ^{o}C for 2 hours in a 100 ml Erlenmeyer flask, with shaking to deplete the cells of intracellular Pi. After this secondary incubation, cells were harvested into 50 mL Falcon tubes by centrifugation at 4,500 g for 15 min at 4 ^{o}C, then resuspended with 1 mL of MOPS-P0 and transferred into a 1.5 mL centrifuge tube. The cells were washed twice with 1mL of MOPS-P0 and then diluted with MOPS-P0 to a variable final OD_{600nm}, depending on the experiment. The cells were kept on ice for no more than 4 hours before conducting the transport experiment.

Nitrocellulose membranes (HAWP02500; EMD Millipore, Co.) were pre-soaked in 1M phosphate buffer (KH_2PO_4 / K_2HPO_4) pH 7.0 for 30 min prior to use. The transport reaction consisted of 20µL of cells added to 360µL of MOPS-P0 and incubated in a water

bath at 30ºC (*S. meliloti*) or 37 ºC (*E. coli*) for 5 minutes before addition of 20 µL of [P^{33}]-labeled inorganic phosphate (150 Ci/mol; Perkin Elmer, Inc.) to the required concentration. Soaked membranes were placed onto a vacuum manifold and 100 µL of the transport reaction was spotted onto the membrane and drained of liquid. Immediately, 5-10 mL of MOPS wash buffer (MOPS-P0 without carbon source) was added to remove extracellular phosphate. After evacuation, membranes were dried using a heat lamp and transferred into a 20 mL glass liquid scintillation vial (Duran, Wheaton, Kimble, Inc.) and 5 mL of Optiphase HiSafe 3 scintillation fluid (#1200.437, Perkin Elmer, Inc.) was added. The samples were counted in a Tri-Carb 2900TR Scintillation Counter (Perkin Elmer, Inc.). Scintillation counts were corrected with toluene treated cells of the same OD_{600nm} (20 % toluene per cells).

To determine the total protein within the transport reactions, cells were centrifuged and resuspend with 1/2 the volume of 1 M NaOH and lysed by boiling at 95 ºC for ~10-15 minutes, then neutralized with 1/2 the volume of 1 M HCl. The protein content of 5-15 µL of the cell lysate was then quantitated using the Bradford method (reference) using Coomassie brilliant blue R-250 solution (Bio-Rad Laboratories, Ltd.). A protein standard curve using bovine serum albumin (BSA) was performed each time, with an R^2 of at least 95 %. All protein quantitation was conducted in triplicates reactions.

2.10.1 - Determination of initial uptake velocity

Strains were tested in biological triplicates and aliquots were sampled at 30, 60 and 90 seconds after addition of [^{33}P]-labeled phosphate. A final cell density of OD_{600nm}

~0.75 (*S. meliloti*) and ~0.5 *(E. coli*) was used, such that no more than 10 % of the total phosphate was taken into cells. A transport reaction concentration of 10 µM Pi (50 Ci/mol) was used, which is ~ 5-fold greater than the reported ~2 µM K_M of *S. meliloti* Pap-Pit system (Voegele, *et al.*, 1997). Initial uptake velocity was determined by the slope of the phosphate accumulation (nmol Pi mg protein^{-1}) over time (min^{-1}). For accurate comparisons, all transport experiments were conducted with the same [^{33}P]-labeled phosphate solution.

2.10.2 - Accumulation and exchange of phosphate in *E. coli*

For accumulation experiments, a 1000 µL reaction (900 µL MOPS-P0, 50 µL of cells OD_{600nm} ~10.0 and 50 µL of [P^{33}]-labeled phosphate) with a final concentration of 500 µM phosphate (2 Ci/mol) was used. All experiments were conducted in biological duplicates with 100µL aliquots of the transport reaction were sampled at various time points (between 1 min and 60 min, depending on the experiment). Cells were mixed by gentle vortexing every 5 minutes.

For exchange experiments, 10 µL of unlabeled phosphate (1 M) was added, which was ~30 times the total phosphate in the reaction. Complementary accumulation and exchange transport reactions were conducted in biological duplicates at the same time, so that a comparison could be made between the two conditions could be tested.

2.10.3 - Enzyme kinetics of phosphate uptake

The kinetics of phosphate uptake of via the Pap-Pit system was determined by measuring the initial velocities of each strain with phosphate concentrations varying from 0.25 µM to 500 µM, depending on the strain. Uptake was measured for duplicate samples at 30, 60 and 90 seconds after addition of [^{33}P]-labeled phosphate. Transport rates were fitted to the Michaelis-Menten equation by non-linear regression function nls() in R using all available data points.

2.11 - Bioinformatic analyses of Pap-Pit

2.11.1 - Identification of Pap and Pit orthologues by profile Hidden Markov Model motif search

A profile Hidden Markov Model (HMM) was used to identify Pfam domains in protein sequences. HMMER software was downloaded from http://hmmer.org/. Release 30.0 of the Pfam database was obtained from the site maintained by EMBL-EBI, http://pfam.xfam.org/. The HMMER 3.0 program hmmbuild was used with Pfam seed alignments of protein families PHO4 (PF01384), PhoU (PF01895) and PhoU_div (PF01865) to generate HMM profiles. Genbank (.gbk) and annotated protein (.faa) files were obtained by ftp through the appropriate NCBI Genome Project site (http://www.ncbi.nlm.nih.gov/). A diverse selection of proteomes were sampled from bacteria (N=1980) and archaea (N=200). A full list of genomes sampled is included in Appendix Table 2. Protein sequences were identified using hmmsearch within the

proteomes querying with the respective HMM profiles. The sequences were then manually curated and refined by HMM best hit (Pfam domain with lowest E-value) and domain organization (additional domains that increase the size of the protein).

2.11.2 - Phylogenetics of Pap and Pit proteins

Protein sequences were aligned by ClustalW (v1.82) and maximum likelihood phylogenies were generated using the RAxML-HPC BlackBox (8.2.9) tools on the CIPRES Science Gateway v3.3 (http://www.phylo.org/). Phylogenies were visualized and annotated based on characteristics of Pap and Pit proteins using FigTree v1.4 9 (http://tree.bio.ed.ac.uk/software/figtree/). Conserved amino acid sequence logos of the profile HMMs were generated using the Skylign web server (http://www.skylign.org/).

2.11.3 - Structural prediction and analysis of *S. meliloti* Pap

I-TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) was used to predict the protein structure, related crystal structures and functional motifs of the *S. meliloti* Pap protein. Protein-protein structural alignments were conducted using TM-ALIGN (https://zhanglab.ccmb.med.umich.edu/TM-align/). Further analysis and visualization of the protein structures were conducted using UCSF Chimera v1.11.2 (https://www.cgl.ucsf.edu/chimera/).

2.11.4 - Membrane topology of Pit proteins

Transmembrane helices of Pit orthologues were predicted using TMHMM v2.0 web server (www.cbs.dtu.dk/services/TMHMM/).

2.11.5 – Profile Hidden-Markov Model (HMM) Logo Visualization

HMM logos of PHO4, PhoU and PhoU_div (Pap) protein domains was visualized using Skylign web server (http://www.skylign.org) (Wheeler *et al.*, 2014). Logos were generated retaining alignment positions that were not empty and with information above background level (using the settings on the web server).

Chapter 3 – Results

3.1 Bioinformatic analyses

3.1.1 - Pap-Pit orthologues are prevalent among bacteria and archaea

To identify orthologues of Pap and Pit, we searched for PhoU_div and PHO4 Pfam domains in 2180 Bacterial (N=1980) and Archaeal (N=200) proteomes, and subsequently curated by domain organization, retaining proteins with PhoU_div or PHO4 motifs as the highest scoring hit. Within these proteomes, a total of 1079 PhoU_div hits (in 1008 proteomes) and 4406 PHO4 hits (in 3122 proteomes) were identified. A full list of Pap, Pit and PhoU protein is included in Appendix Data Table 3.

Using the annotated start/stop annotations, 874 PhoU_div/PHO4 pairs were identified with ORFs with a distance of less than 200 base pairs and were denoted as putative Pap-Pit orthologues. Paired orthologues were identified in 813 proteomes (32% of proteomes containing PHO4 hits), indicating that this system is present in a significant fraction of prokaryotic organisms. Furthermore, *pap-pit* genes were found to be encoded as co-directional *pap-pit* genes in both Bacteria (99.5%) and Archaea (91%). Only 9 convergently and 5 divergently transcribed *pap-pit* orthologues were identified (genes transcribed towards each other or away from each other, respectively).

The phylogenetic distribution of Pap-Pit was found to be highly diverse, as putative orthologues were present in most phyla except those with few representatives (Table 3-1). Pap-Pit orthologues were not identified in the archaeal phyla Korarchaeota,

Lokiarchaeota, Nanoarchaeota and Nanohaloarchaeota, all of which have few sequenced genomes.

54 genomes were found to have multiple orthologues of *pap-pit*, of which 4 were identified in *Accumulibacter phosphatis*, two of which are separated by a single unannotated ORF and another is a Pap-PitA-like operon.

3.1.2 - Phylogenetics reveal two distinct clades of Pap-Pit, distinct from *E. coli* PitA

SmPap and SmPit encode proteins of 214 and 334 amino acids, respectively. An analysis of the size distribution of Pap and Pit proteins indicates that there are two clusters of (Pap)-(Pit): the major cluster with of (210)-(335) a.a. and a minor group of (223)-(417) a.a. (Figure 3-2A). This indicates that within the category of Pap-Pit proteins, there are two sub-groups of different sized proteins. Additionally, a maximum likelihood phylogeny of concatenated Pap-Pit proteins forms similar clusters that correspond to Pit protein size, but the two groups are not entirely monophyletic indicating the phylogenetic distinctions are not due solely to protein size (Figure 3-2B).

467 PhoU_div hits not paired with an adjacent PHO4 hit were identified. This is not surprising, as there may be mutations (recombination resulting in deletion, inversions, etc.) abolishing or segregating the cognate gene pairs. Alternatively, the lack of an adjacent PHO4 hit may be due to sequencing or annotation errors that prevent detection of either protein.

Similarly, 1154 unpaired Pit (herein called "Pit-only") proteins were also identified with PHO4 as a best hit, but not adjacent to a PhoU_div hit. This classification includes *E. coli* PitA/PitB but also *S. meliloti* Pit2. However, inspection of the protein size versus E-value show that the size distribution of Pit-only proteins varied greatly (Figure 3-4A). These proteins were on average 461 ± 29 a.a. in size. However there is a cluster of 77 Pit-only proteins with a mean size of 753 a.a. and E-value E-34, which indicate these are true hits to large proteins that have a PHO4 domain. There were also several smaller clusters between 300-500 a.a. in size with variable E-values between E-30 to E-120. Proteins of less than 300 a.a. in size were excluded from alignment, as they are likely truncated proteins.

Interestingly, 142 of the 1154 Pit-only proteins were identified among Gammaproteobacteria-Enterobacteria, in contrast to only 7 *pap-pit* orthologues, even though this phylogenetic class represented 125 of 2180 proteomes in the query. This group included proteins with a mean of 507 ± 42 a.a in size, which includes *E. coli* PitA.

A histogram of *pap*-associated Pit proteins compared to Pit-only reflect the differences between SmPit and EcPitA, as all *pap*-associated Pit proteins have a mean length less than 450 a.a. (Figure 3-4B).

Phylogeny of all Pit proteins (Pap-associated Pit and Pit-only) reflect the alignment between Pap-associated Pit proteins (Figure 3-5). The two clusters of Pap-Pit proteins remain as distinct clades, with Pit-only proteins interceding in between. Interestingly, *E. coli* PitA-like proteins were found to formed a monophyletic clade of 327 proteins, of which >65% are Enterobacteria (Figure 3-5).

Including eukaryotic Pit proteins (*N. crassa* PHO-4, *S. cerevisciae* Pho89, and *H. sapiens* Pit1) as outgroups did not alter the topology of the phylogeny, which may suggest that some bacterial Pit proteins are more distantly related to each other than eukaryotic Pit proteins.

HMM logos of *pap*-associated Pit and PitA-like proteins revealed that both consist of two PHO4 domains separated by a region of low conservation (Figure 3-7). PHO4 domains consist of two PHO4 signature sequence motifs like that of eukaryotic Pit proteins (Bottger *et al.*, 2011). The *pap*-associated Pit and PitA-like proteins have distinct PHO4 signature sequence motif (as shown in the comparison of the N- and C-terminal PHO4 domains). PitA-like proteins have an N-terminal PHO4 motif with two highly conserved tryptophan (W) residues that are also present in *pap*-associated Pit proteins albeit less conserved (Figure 3-7A). Additionally, the C-terminal PHO4 domain of PitA-like proteins has a large non-conserved insertion between the signature sequences (Figure 3-7B). A representative alignment with putative transmembrane helices (TMH) prediction of these two groups (using SmPit and EcPitA) show that there is a large gap region of approximately 120 a.a. between helices 6 and 7 which forms a cytosolic loop (Figure 3-8). SmPit and EcPitA are also both predicted to have 10 TMH, with both C- and N-termini in the periplasm.

3.1.3 - Pap proteins are similar to PhoU

The 863 Pap and 2028 PhoU proteins were used to generate HMM logos of the two families of proteins. The HMM logo of PhoU and Pap proteins were found to be

quite similar: both shared conserved PhoU-like metal-binding motifs (E/DXXXD) and a highly conserved intramotif region (Figure 3-9). The intramotif regions of Pap (PXXRXD) differed from PhoU (PXXXXDLR), however they both share a highly conserved proline, arginine and aspartate residues. Additionally, the PhoU motifs of Pap proteins are less conserved compared to those of PhoU proteins. An alignment of PhoU and Pap proteins (Figure 3-10) shows that Pap and PhoU proteins not only share similar conserved residues, but also

Several crystal structures of PhoU proteins from bacteria have been determined and published (Liu *et al*., 2005; Oganesyan *et al*., 2005). These structures show that PhoU proteins form a six parallel alpha helices and conform into dimeric or trimeric complexes (Figure 3-11). Two crystal structures of Pap-like orthologues have been determined: *Bacteroides thetaiotaomicron* (BtPap; PDB: 3L39) and *Shewanella oneidensis* (SoPap; PDB: 2IIU), and both structures indicate Pap-like proteins form a bundle of six parallel alpha helices, predicted to form a trimeric structure (Figure 3-11B and 3-11C).

I-TASSER structural prediction of SmPap generated a model that forms the canonical PhoU/Pap-like crystal structure of six parallel α-helices (Figure 3-12A). The SmPap PhoU-motifs are also predicted to bind to Fe^{+2} or PO_4^{-3}, like PhoU proteins which are believed to have metal-binding motifs (Liu *et al.,* 2005) (Figure 3-12B). The PhoU-motifs of α2-α3, and α5-α6 all face one side of the protein referred to as the ‘inner face’ of the Pap protein, whereas α1 and α4 form the ‘outer face’ and are less conserved. An electrostatic potential plot of SmPap proteins indicate that the inner face is highly negatively charged compared to the outer face which is neutral (Figure 3-12C).

Interestingly, a negatively charged electrostatic potential is primarily found on the interior face of SmPap that is predicted to form the core of the homotrimeric complex (Figure 3-13). This may be indicative of possible interactions with positively charged molecules (such as metal ions) or with positively-charged residues of other proteins.

3.1.4 - Alternative proteins containing PhoU and PhoU_div domain hits

When examining proteins with PhoU_div hits but not as the best hit (in terms of E-value), various proteins were identified (Table 3-14A). It should be noted that with some exceptions, most PhoU_div hit E-values were greater than E-5. Although, proteins with a major-facilitator super family (MFS-1) domain as the best hit had a mean E-value of < E-17, suggesting that these proteins likely contain a real PhoU_div domain. Additionally, sugar transporter, Na+/Pi cotransporter, and TrkA_C domains were most commonly identified in proteins that also had a PhoU_div hit.

Several proteins with PhoU hits were also found to have a best-hit domains other than PhoU (Table 3-14B). As expected, these included a considerable number of proteins containing PhoU_div domains as a best hit, which is likely due to cross-identification of the two protein domains. However, there were a large number of hits for other domains, including Na_Pi_cotrans, TrkA_C and MazE-antitoxin.

3.2 - Evidence that *S. meliloti* Pap is required for Pit-mediated Phosphate Uptake

3.2.1 - *S. meliloti pap* and *pit* are required for phosphate uptake via Pit

To study the effects of Δ*pap*, Δ*pit* and Δ*pap-pit* deletions on Pit-dependent Pi uptake, a parent strain was used where expression of the Pho regulon (PhoB, PhoR, as and PstSCAB-PhoU system) is IPTG-dependent (diCenzo *et al.*, 2017). Routine culturing of *S. meliloti* with the deletion mutations in the presence of IPTG permitted growth without inadvertent selection of Pi uptake revertants. Growth in MOPS minimal medium with 2 mM Pi (MOPS-P2) as the sole source of P and in the absence of IPTG (where the Pho regulon is not induced) is suggestive of Pit-dependent Pi uptake.

The growth phenotype of RmP3924 (Δ*pap*) was not complemented by integration of pTH3192 (promoter-*pap*) as no growth or Pi uptake was observed (Figure 3-15A and 3-15B). However, pTH3193 (promoter-*pap-pit*^{truncated}) was mated into RmP3924 resulting in two possible recombinants: insertion of the *pap* gene directly upstream of the native *pit* gene or insertion of *pap-pit*^{truncated} under a separate promoter (Figure 3-16A). Both recombinants were independently isolated and confirmed by colony PCR (Figure 3-16B). Complementation was observed when the *pap-pit* operon was restored, and not when the genes were independently expressed (Figure 3-16C).

Integration of pTH3194 (*pap*^{truncated}*-pit*) at the Δ*pit* locus resulted in insertion of the *pit* gene directly downstream of the native *pap* gene and complementation of the RmP3925 (Δ*pit*) mutant phenotype (Figure 3-17A). Growth in MOPS-P2 and Pi uptake of RmP3925 was not restored by pTH3195 (promoter-*pit*) (Figure 3-17B and 3-17C).

This suggests that expression of *pit* from a Δ*pap* mutant or via the promoter-*pit* construct produces insufficient levels of the Pit protein, and is not due to secondary mutations due to a lack of the *pap* gene.

In contrast, RmP3926 (Δ*pap-pit*) was not complemented with either pTH3193 or pTH3194, which indicates that neither *pap* nor *pit* genes, are sufficient for Pi uptake alone. This evidence may also suggest that deletion of the *pap* gene may have polar effects on expression of the downstream *pit* gene, even though the ribosome binding site of the *pap* gene is retained and there are no other start codons other than the native *pit* ATG.

3.2.2 - Detection of recombinant Pit~tag fusion proteins

The complementation experiments support the hypothesis that the *S. meliloti* Pit protein is insufficient for Pi uptake alone, however, evidence that the Pit protein is translated is lacking. To detect the Pit protein, various tags were fused to the C-terminus of the Pit protein and it was expressed as a wildtype *pap-pit* operon in pUCP30T (pTH3153, His10; pTH3198, *phoA/lacZ* reporter fusion, pTH3200, bovine rhodopsin 1D4 peptide; pTH3201, FLAG; pTH3202 Strep-II;).

The Pap-Pit~PhoA/LacZ translational reporter fusion construct cloned into pUCP30T (pTH3198) was expressed from the native *pap-pit* promoter. The *S. meliloti* RmP3926 (Δ*pap-pit*) (pTH3198 - promoter *pap-pit~phoA/lacZ*) recombinant strain appeared as blue colonies on LBmc agar with 40 µg/mL X-phos and without IPTG, demonstrating that the recombinant Pap-Pit~PhoA/LacZ system can complement the growth and Pi uptake phenotypes of the Δ*pap-pit* mutant (Figure 3-18). RmP3926

(pTH2824 – wildtype *pap-pit*) recombinant cells had slightly blue colonies on the same media, indicating that the greater AP activity in RmP3926 (pTH3198) was due to the recombinant Pit~PhoA/LacZ protein. However, RmP3926 (pTH3196 - promoter *pit~phoA/lacZ*) was unable to grow on LBmc agar without IPTG and was just as blue as RmP3926 (pTH2824 – wildtype *pap-pit*) on LBmc agar with IPTG and X-phos.

RmP3926 (Δ*pap-pit*) with *pap-pit* or *pap-pit~phoA/lacZ* showed no detectable difference in AP activity when cells were incubated in MOPS minimal medium (20 mM glucose) with 0 mM Pi (MOPS-P0) or 2 mM Pi (MOPS-P2) (Figure 3-19A). An in-gel zymogram of AP activity was conducted to detect reporter expression of Pit~PhoA/LacZ, after protein separation of soluble and insoluble fractions in a non-denaturing PAGE (Figure 3-19B). The cytoplasmic fractions of wildtype Pap-Pit and Pap-Pit~PhoA/LacZ, produced AP activity at the bottom of the gel, which likely corresponds to the *S. meliloti* PhoX that was previously characterized as a periplasmic AP under PhoB-dependent expression induced in low Pi environments (Zaheer *et al.,* 2009). However, the insoluble fraction of RmP3926 (pTH3198) also showed elevated levels of AP activity at the top of the gel, which could correspond to the Pit~PhoA/LacZ fusion protein. However, the same AP activity at the top of the gel was not observed for RmP3926 (pTH3196). This indicates that the Pit~PhoA/LacZ recombinant protein is only efficiently expressed from a wildtype *pap-pit* operon and not when the *pap* gene has been removed completely. As this observation was made from a non-denaturing PAGE, the expected size of the recombinant Pit~PhoA/LacZ protein could not be determined, particularly since the Pit protein would form interaction with many different hydrophobic membrane particles.

Detection of the recombinant Pit~His10 was conducted by SDS-PAGE and western blot using rabbit monoclonal antibodies against the polyhistidine tag. From RmP3926 (*pap-pit*~His10), the recombinant Pit~His10 was not detectable in cell lysates, insoluble or soluble fractions, in comparison with wildtype *pap-pit* (Figure 3-20). The same observation was made even when using various detergents such as 0.5% SDS or Triton X-100, as well as 2% NP-40 during cell lysis. Similar observations were made for RmP3926 (pTH3200 - *pap-pit*~Rho1D4) using mouse monoclonal anti-1D4 antibodies (Appendix Figure 4).

3.2.2 - Pap alanine-substitution mutants are Pit phosphate uptake defective

The pUCP30T-derivatives which contained wildtype *pap-pit* with single missense mutations in Pap were mated into RmP3926 (Δ*pap-pit*) to determine whether these mutations had effects on Pi uptake via Pit (shown in Table 2-4 and Figure 3-10). Most recombinants were found to have moderate or wildtype rates of growth in MOPS-P2, except for the PapP75A and PapR78A mutants (Figure 3-21A). Similarly, no Pi uptake was observed for these strains in a transport assay with 10 µM Pi (Figure 3-21B).

3.2.3 - *S. meliloti* Pap protein-protein interactions

Previously, *in vivo* formaldehyde cross-linking of *S. meliloti* PhoU as well as PhoU-PhoU bacterial two-hybrid interactions demonstrated that PhoU proteins form homo-oligomers, likely homodimers (diCenzo *et al.*, 2017). Similarly, detection of Pap protein-protein interactions by *in vivo* formaldehyde cross-linking produced a novel band

of ~50kDa (Figure 3-22). This novel band was present only in cell lysates from formaldehyde-treated cells of wildtype RmP110 but not RmP1628 (Δ*pap*), which indicates that this novel band is due to cross-linking of the Pap protein and not a protein detected by non-specific interaction with the anti-Pap antisera. This band is still present in RmP1629 (Δ*pit*), which indicates that this interaction is not between Pap-Pit and is not influenced by Pit (estimated 23kDa and 34kDa in size).

Additionally, a bacterial two-hybrid (BACTH) system was employed to test the interactions between Pap-Pap as well as Pap-Pit proteins by creating recombinant fusions with the T18 and T25 fragments of the bacterial adenylate cyclase enzyme (Karimova *et al.*, 1998). A significant interaction was observed for Pap-Pap interactions when the cells were grown in LB and high Pi media, however there was no significant interactions detected between Pap and Pit proteins, regardless of the whether N- or C-terminal fusions to the T18 and T25 fragments to the proteins were examined (Appendix Table 5).

3.3 - Heterologous expression of Pap-Pit orthologues in Pi-uptake deficient *E. coli* MT2006

The *pap-pit* operons and the *pit* genes from *S. meliloti* (SmPapPit), *B. thetaiotaomicron* (BtPapPit) and *S. oneidensis* (SoPapPit) were amplified and cloned into pMW119, where their expression was driven from the *lac* promoter. These systems were then tested for Pi uptake by growth complementation of *E. coli* MT2006, where growth with Pi as the sole source of P would be indicative of Pi uptake (Motomura et al., 2011). Additionally, MT2006 with pMW119 carrying *E. coli pitA* (EcPitA) was used as a

positive control for physiologically normal levels of Pi transport from a plasmid when *E. coli* is grown in high Pi environments. In addition to growth experiments, Pi uptake assays were conducted to confirm the results from growth experiments and the Pi uptake kinetics of were examined. The effects of the previously characterized amino acid substitutions of the *S. meliloti* Pap protein were also examined to determine whether the phenotypes observed in *S. meliloti* would also reflect those observed when expressed in *E. coli.*

3.3.1 - Complementation of MT2006 on M9 minimal media

Growth of *E. coli* MT2006 (BtPapPit) and MT2006 (SoPapPit) in M9-mannitol was similar to MT2006 (EcPitA) (Figure 3-23). However, growth of *E. coli* MT2006 (pMW119), MT2006 (SmPapPit), and MT2006 (BtPit) was not observed in M9-mannitol unless glycerol-3-phosphate (G3P) was added to the media (Figure 3-23). This suggests that MT2006 (SmPapPit) or MT2006 (BtPit) did not grow due to P starvation. Interestingly, MT2006 (SoPit) reached the same stationary phase density 5 hours later than MT2006 (SoPapPit) in M9-IPTG, which demonstrates that SoPit has some Pi uptake even in the absence of SoPap, but only when over-expressed.

Pi uptake rates were also measured using by conducting transport assays and the data from these assays were found to be consistent that of the observed growth data (i.e. if no growth was observed, no Pi uptake were also observed and vice versa) (Table 3-25). The initial rate of Pi uptake of MT2006 (BtPapPit) and MT2006 (SoPapPit) was approximately 2-fold greater than that of MT2006 (EcPitA) (Table 3-24). Insignificant

differences in Pi uptake were observed between MT2006 (pMW119) and MT2006 (SmPit) and (BtPit) even with cultures containing the inducer IPTG. In contrast, Pi uptake by MT2006 (SoPit) was observed when IPTG was present in the media with a transport velocity found to be ~ 1/5 of the initial uptake velocity of MT2006 (SoPapPit).

3.3.2 - Pap-Pit systems have greater affinity to Pi than *E. coli* PitA

Pi uptake kinetics of the various Pit systems were tested by measuring the uptake velocity in reactions with Pi concentrations between 0.25 and 500µM (Figure and Table 3-25). Saturation kinetics were observed for the MT2006 carrying BtPapPit and SoPapPit, with observed K_M values of 4.1-8.5µM and 15µM, respectively, which were more than 2-fold lower than that of EcPitA (approximately 27-37µM). These observations reflect the previously observed K_M of SmPapPit (1-2 µM Pi) which indicate that Pap-Pit systems have greater affinity for Pi than EcPitA (Voegele *et al.*, 1997).

Saturating Pi uptake was not observed for MT2006 (SoPit), even at 500µM Pi. A linearly increasing velocity was observed with increased Pi concentrations, which may indicate that the SoPit protein, in the absence of the Pap protein, functions as a passive transport system with a high K_M.

3.3.3 - Pap-Pit systems function as uptake/exchange Pi systems in E. coli

Previous experiments with *E. coli pst* mutants revealed that the *E. coli* PitA system had Pi export capacities, as radioactive Pi was observed to leave the cells when excess unlabeled Pi was added to the transport reaction after several minutes of

accumulation (Rosenberg *et al*., 1982). Similarly, our data shows a linear Pi accumulation over a 60-minute period, and exchange of Pi occurred when 30-fold excess unlabeled Pi was added to the reaction (Figure 3-26). This exchange was observed for BtPapPit, SoPapPit, EcPitA and also SoPit. Additionally, the rate of Pi exchange (rate of labeled Pi leaving the cells) was approximately equal to the rate of Pi accumulation, which indicates that the rate of uptake observed after 10-60 minutes is the net uptake of Pi due to positive influx:efflux via the Pit systems.

3.3.4 - *S. meliloti* Pap mutants and Pit uptake phenotypes

The Sm *pap-pit* genes carrying *pap* mutations were subcloned from pUCP30T into the high copy-number vector, pUC119, carrying the *lac* promoter. These vectors were then transformed into *E. coli* MT2006 to determine whether these mutations affected Pi uptake via this system.

Growth experiments demonstrated that alanine substitution mutations of conserved residues resulted in complete loss of growth in M9 but not in M9+G3P, which indicate that these residues are essential for Pi uptake via SmPapPit in *E. coli* (Figure 3-27). These residues were previously identified (see Figure 3-1) to be highly conserved among Pap orthologues, which may indicate they have some significant role in the function of the Pap protein. Additional mutations resulted in a reduced growth velocity or maximal density, which suggest these residues may be functionally involved but are not essential in the Pi transport process.

Interestingly, the transport phenotypes were not reflective of the observed growth phenotypes: PapD58A, PapD93A, and PapD90A/H96Y all grew in M9-mannitol media, but no Pi uptake was observed (Table 3-28). PapD90A, PapE193A and PapD197A grew poorly on M9-mannitol but had observable Pi uptake rates like wildtype SmPapPit, which may indicate that the mutations may affect affinity for Pi but not the actual rate of Pi uptake. The incongruences between growth and Pi uptake results may reflect the fact that the M9-mannitol media and the transport reaction contained different concentrations of Pi (transport assays were conducted with 10 μM Pi, whereas M9 contains approximately 69 mM Pi). This could have significant consequences if mutations affected the K_M of the SmPapPit system.

Intriguingly, Pap proteins of different sizes were observed for anti-Pap western blot of cell lysates of these mutants (Figure 3-29). PapE54A, PapD58A, and D93A had a Pap protein band which was different from that of the wildtype Pap protein. However, PapD160A lacked any observable Pap protein at all, which is strange as the PapD160A mutant was observed to grow in M9-mannitol and had observable Pi uptake.

3.3.5 – Complementation of *Shewanella* and *Bacteroides pit*

Because *E. coli* MT2006 carrying SoPit alone was observed to have some growth on M9 20mM mannitol without G3P and had observable Pi uptake above background, the reduced phenotype may be rescued with reintroduction of the SoPap protein. Complementation of MT2006 was conducted by co-transforming the cells with pMW119 carrying SoPit and pTH3232 (Sp^R variant of pTH1937) carrying SoPap. However,

complementation of growth in M9-mannitol with IPTG was not observed (Figure 3-30A) as the growth of MT2006 carrying SoPap and SoPit grew slower than MT2006 carrying SoPit alone. Additionally, a parallel experiment was conducted with BtPap and BtPit where no complementation was observed (Figure 3-30B).

Chapter 4 – Discussion

4.1 - Pap is required for active Pi uptake via the Pit transporter

Phosphorus (P) is present in several core biomolecules including DNA, RNA and phospholipids. Thus, in media containing inorganic phosphate (Pi) as the sole source of phosphorus, cells require a functional Pi transport system for growth. In this work, Pap was demonstrated to be involved in Pi uptake via Pit, as Pi uptake and growth in minimal media containing Pi as the sole source of phosphorus were observed only when the *pap* and *pit* genes were present as an operon. This was the case for *S. meliloti pap-pit,* as no Pi uptake or growth was observed when the cognate genes were present under separate promoters or when only the *pit* gene was present.

Anti-Pap western blots showed that SmPap was present at similar concentrations when expressed from the native promoter, regardless of whether the cognate *pit* gene was present (Appendix Figure 6). However, recombinant Pit transporter was not detectable by Western blot using anti-His or anti-Rho1D4 antibodies, when expressed from the wildtype *pap-pit* operon (Figure 3-20). Several attempts were made to solubilize the bacterial membrane fraction during lysis with various detergents (0.5% Tween-20, Triton X-100, SDS, 2% NP-40) prior to SDS-PAGE in order to improve the detection of SmPit~His10 using anti-His antibodies, however no noticeable signal was detected. In a native PAGE followed by an in-gel zymogram of AP activity, the recombinant Pit~PhoA/LacZ protein may have been detected from *pap-pit~phoA/lacZ* but not *pit~phoA/lacZ* (Figure 3-19B). However, without significant size-separation through the

native PAGE, it is unclear whether this was the recombinant Pit protein. Therefore, it is inconclusive whether the SmPit alone is insufficient for Pi uptake in *S. meliloti*, and further experiments are required to determine whether this is the case.

It should also be noted that here, the SmPit protein was examined when expressed from a native promoter in *S. meliloti* on the chromosome. As the *S. meliloti pap-pit* genes are transcribed at a moderate to low level, this moderate expression only adds to the difficulty of Pit detection (Supplemental Figure S7). Thus, it may be necessary to screen for the Pit protein by over-expressing and purifying different recombinant Pit~tag proteins using a variety of detergents followed by anti-tag western blot. This can be done by cloning the *pap-pit~tag* into pTH3227 (pUCP30T with the *S. meliloti pcaD* promoter), which carries a protocatechuic acid inducible promoter to drive expression of the *pap-pit*~tag operon (White *et al.*, 2012).

However, detection of SmPit from promoter-*pit* (scar-less deletion of *pap*) may not be straight-forward, as the transcript or protein may be unstable in the absence of *pap*. For example, if Pap functions as a protein chaperone involved in protein-folding, structural maintenance or membrane localization of Pit, the absence of Pap would likely result in accumulation of the hydrophobic Pit polypeptide and followed by degradation (Baneyx & Mujacic, 2004; Dalbey *et al*., 2011). Alternatively, transcription and translation of *pit* may be inefficient due polar effects caused by the deletion of the *pap* coding sequence. It may be that the *pap-pit* polycistronic mRNA is translationally coupled (the two open-reading frames are translated successively by the same ribosome), which has been observed for other polycistronic mRNAs where the transcripts overlap

(Wang *et al.*, 2014). The exact benefit is not clear, but there is some evidence that translational coupling reduces variability in translational efficiency and ensures similar molecular ratios of proteins that are involved in a specific function (Okuda *et al.*, 2007). There could also be specific interactions between the upstream and downstream genes, where the former gene encodes a chaperone required for processing, folding, or oligomerization the latter protein (Wang *et al.*, 2014). Whether the *S. meliloti pap-pit* genes are translationally coupled has not been directly investigated. Translational coupling between the *pap-pit* reading frames could be explored by generating non-sense mutations (creating a premature stop codon) towards the region 3' end of the *pap* gene. These pre-mature stop codons would result in a reduction or prevention of *pit* translation (van de Guchte *et al.*, 1991).

Alternatively, *S. meliloti* expressing *pap-pit* carrying missense mutations of key conserved residues in Pap (particularly P75A and R78A), were shown to have defective growth and reduced Pi uptake, even though the *pit* gene and the promoter regions were unchanged (Figure 3-21 and Figure 3-28, Figure 3-29). It is unlikely that these 1-2 bp changes would result in polar effects on the downstream *pit* gene, and it may be interesting to examine whether the SmPit protein is efficiently translated in these mutants. If the SmPit protein is present but the SmPap mutant results in defective Pi uptake, the SmPap protein likely has some direct involvement in Pi uptake. If the SmPit protein is not present and the SmPap mutant results in defective Pi uptake, it may be that the SmPap protein has chaperone-like functions as mentioned above.

4.2 - Characteristics of Heterologous Pap-Pit systems expressed in *E. coli*

Pi uptake-deficient *E. coli* MT2006 was also used for testing heterologous expression of Pap-Pit systems. As expected, expression of the BtPapPit, SoPapPit, and EcPitA permitted growth of MT2006 on M9-mannitol (Figures 3-21). BtPapPit and SoPapPit both were found to have apparent K_M values (17 and 8.5 µM, respectively) that were 2-3 fold less than that of EcPitA (27 µM). However, all the Pap-Pit and PitA systems shared similar V_{max} values of 400-500 nmol Pi/mg protein/min.

These values are similar to the previously reported K_M of 1-2 µM but different from the V_{max} of 0.6 nmol/Pi/mg protein/min measured for *S. meliloti* RmG490 (*pstC and phoC* mutants) with only wildtype *pap-pit* (Voegele et al., 1997). The differences in V_{max} may be because the Pit systems in *E. coli* are expressed from a low copy-number plasmid (pMW119 is present in 1-10 copies) and driven by the *lac* promoter, whereas in *S. meliloti* RmG490, *pap-pit* was present in single copy on the chromosome under the native promoter.

Exchange of Pi was observed for the Pit systems when excess unlabeled Pi was added after accumulation of radiolabeled Pi, which suggests that the observed Pi uptake is due to the net uptake of Pi (the difference between the v_{import} -v_{export} of Pi). The observed net uptake of Pi is like due to assimilation and incorporation of the intracellular Pi into biomolecules as a result of Pi esterification, including ATP synthesis and other reactions (Rosenberg *et al.*, 1982). These results suggest that the Pap-Pit and PitA-like transport systems catalyze similar reversible transport reactions with similar characteristics that

serve similar physiological functions, such as maintaining cytosolic Pi homeostasis in high Pi environments by permitting the exchange of Pi.

4.3 - SoPit alone as a Pi uptake system

The IPTG-induced expression of SoPit in *E. coli* MT2006 resulted in moderate growth in M9-mannitol relative to the MT2006 with the empty pMW119 vector alone (Figure 3-23). The IPTG-induced expression of SoPit also resulted Pi uptake with a greatly reduced K_M and uptake velocities relative to *E. coli* MT2006 carrying SoPapPit (Figure 3-25). The precise K_M and V_{max} of SoPit could not be determined, as the rate of Pi uptake from the system was not observed to saturate with increasing concentrations of Pi. Rather, the observed rate of Pi uptake increased linearly with increasing concentrations of Pi. The actual K_M of SoPit is likely in the mM range, similar to what is observed for *S. cerevisiae* Pi uptake systems (Wykoff & O'Shea, 2001). Attempts to calculate the SoPit K_M by subtracting background uptake by MT2006 (pMW119) were not also unsuccessful as the observed background rates were highly variable (Appendix Figure 8).

While SoPit alone appears capable of transporting Pi, neither the SmPit or BtPit proteins were observed to transport Pi into *E. coli* or *S. meliloti*. Interestingly, the phylogenetic analyses showed SoPit to be present within the clade of Pit proteins of larger size (~420 a.a.), distinct from the clade that includes SmPit and BtPit (~350 a.a.) (Figure 3-5). An alignment of SmPit, SoPit, BtPit and EcPitA proteins show a 75 a.a. insertion between the two PHO4 domains of SoPit, different from the insertion in EcPitA (Appendix Figure 9).

The Pi uptake and growth observed in cells carrying SoPit in the absence of SoPap is definitive evidence that the SoPit protein is present within the inner membrane of the cell. However, *E. coli* MT2006 with SoPap and SoPit expressed on independently-replicating plasmids does not restore Pi uptake to wildtype levels permitting growth in M9-mannitol with IPTG (Figure 3-30). The lack of total complementation has not yet been investigated, but may be due to poor expression of the genes.

4.4 - Pap-Pit is a diverse prokaryotic Pi uptake system, distinct from PitA-like transporters

One third of bacterial and archaeal genomes contained *pap-pit* orthologues and which demonstrate that Pap-Pit is broadly distributed. The evidence presented here suggests that Pap-Pit is a Pi transporter distinct from the canonical EcPitA-like systems. The phylogeny of Pit orthologues (Figure 3-5) supports this hypothesis, as PitA-like proteins form a distinct clade that is not found with any Pap-associated Pit proteins. *pap*-associated Pit proteins formed two clades (a major and minor clade), fairly consistent with the size-differences of the Pit proteins (350 a.a. compared to 420 a.a.). The phylogeny of the cognate Pap proteins was congruent with that of their associated Pit proteins as both formed the same distinct major and minor clades, suggesting Pap-Pit proteins share functional relationships and co-evolve. However, poor bootstrap values may warrant further refinement of the phylogenetic analysis.

Pap-Pit systems were identified in 36 of 54 prokaryotic classes examined, compared to 27 of 54 prokaryotic classes for Pit-only systems. Interestingly, there were

several classes of which there were a significantly different number of identified Pit systems (Table 3-1). The one of most relevance is Gammaproteobacteria-Enterobacteria, with only 7 Pap-Pit systems identified compared to 142 Pit-only orthologues of which all were found to be PitA-like systems. There may be evolutionary events during the division of Enterobacteria from other bacterial groups that could explain the selective pressure towards PitA-like systems rather than Pap-Pit systems. One such example may be the absence of exposure to lower Pi environments where PitA may be more efficient, as Enterobacteria includes many species that are commonly associated with a host organism (Williams *et al*., 2010). However, Pit-only systems are also predominant within Archaea, where twice the number of Pit orthologues identified compared to Pap-Pit systems. This may suggest that Pit-only proteins were likely the evolutionary origin of Pit proteins. Based on this theory, it may be that Pap-Pit systems evolved after the separation of Enterobacteria from other Bacterial classes, or Pap-Pit was lost from Enterobacteria due to a lack of selective pressure to maintain the system. Further research into the distribution and evolutionary origin of the Pit uptake systems within prokaryotes may provide insight as to the evolutionary and metabolic requirements leading to selection of different Pi transporter systems.

The *pap-pit* orthologues almost always occur as a co-transcribed operon (>95%) which supports the hypothesis that Pap-Pit form a two-protein transport system. Not only does the co-localization of the *pap* and *pit* genes suggest they are functionally related, but the conserved order of genes suggest that Pap-Pit may have an associated efficiency of assembly or interaction, which supports the hypothesis of a modulatory or accessory role

for Pap protein, as translation of Pap would occur prior to that of Pit (Wells *et al.*, 2016). The putative Pap-Pap interactions observed may somehow mediate the interaction of Pit proteins, possibly with membrane localization proteins or between Pit-Pit homodimers, as other Pit proteins are believed to form dimers (Bøttger & Pedersen, 2005; Yan, 2015). Pap and Pit may also be coupled during the translation process, which has also been observed for the bacterial luciferase complex, which may produce a more efficient interaction between the proteins (Shieh *et al.*, 2015).

4.5 - Comparison of PhoU and Pap (PhoU_div)

PhoU and Pap share similar domain organization, protein structure, and mutant phenotype. However, PhoU and Pap proteins were found to be distinct protein families (Appendix Figure 10). Deletion mutations of *S. meliloti* PhoU or Pap result in loss of function phenotypes*: pap* mutations result in lack of Pi uptake via Pit, and PhoU mutations result in constitutive Pi uptake via PstSCAB and constitutive Pho Regulon induction (diCenzo *et al.*, 2017). However, the evolutionary relationship of the proteins suggests that although they share similar features, Pap and PhoU are distinct protein families that form phylogenetic clades (with some small exceptions) (Apendix Figure S10). Pap and PhoU are likely evolved from a common ancestral protein, with divergence due to adaptations that modulate the cognate transport system.

PhoU and PhoU_div protein motifs, although distinct, were commonly misidentified during HMM motif detection, as almost 1/4 of hits shared both HMM motifs, but with different E-values. This is most likely due to the highly conserved PhoU

metal-binding motifs present within each protein (Figure 3-3). Missense mutations of these residues of SmPap (PapE54A, D58A, D93A) resulted in a dramatic decrease in Pi uptake when expressed in both *S. meliloti* and *E. coli* (Figures 3-21 and 3-27). Similarly, a missense mutation of *E. coli* PhoU (E100A/R101A and E200A/R201A double missense mutants) resulted in constitutive PstSCAB activity and increased polyPi accumulation (Gardner *et al.,* 2014). Because of the uncanny similarity between the conserved residues of the metal-binding motifs, they likely share a common molecular mechanism or interaction that is required for the function of these proteins, possibly via a direct interaction with a heterologous protein such as Pit or PstB, respectively, which has been explored by others (Gardner *et al*., 2015; diCenzo *et al*.,2017).

The major amino acid differences between Pap and PhoU are the intramotif residues within the N-terminal PhoU domains of Pap (PXXRXD[I/L]) and PhoU (PXAXDLR) (Figure 3-9). Although these motifs do not appear to be similar, the conserved residues appear to be functionally required in SmPap for Pi uptake (P75A and R78A). The distinct functions of these residues have not been directly explored, however the conserved proline (P) may be required for conformational rigidity of the secondary structure of the PhoU and Pap-like proteins, as they are typically found in kink regions between α-helices and β-sheets (MacArthur and Thornton, 1991).

4.6 - Comparison of Pap-associated Pit and PitA proteins

Pap-Pit systems were found to be more diverse compared to *E. coli* PitA-like systems, but primarily found in Enterobacteria (Figure 3-6). Additionally, we observed a

lower K_M (greater affinity) for Pap-Pit systems compared to EcPitA, although with similar maximum velocities, which may indicate that these systems serve similar physiological functions in maintaining cytoplasmic Pi concentrations. However, the difference in K_M may indicate that PitA-like systems evolved under different conditions to have a lower affinity towards Pi. Most intriguingly, *pap*-associated Pit proteins lack the ~100-120 a.a. intracellular loop region within the C-terminal PHO4 domain that is conserved in PitA-like proteins (Figure 3-7). In comparison, the same intracellular loop region in *pap*-associated Pit proteins is only 5-10 a.a.

The mechanism and function of this loop in Pit proteins is not yet known but is of particular interest to determining function of the Pit proteins, as this loop is also present in eukaryotic Pit proteins as well as other members of the MFS transporters (Weinglass *et al.,* 2000; Salaün *et al.,* 2002). The structure and functions of the intracellular loop region is not known, but there is evidence from the human dopamine transport system, where mutations to specific conserved cysteine residues of the intracellular loop region resulted in reduced K_M and V_{max} (Xu *et al*., 2010). However, the intracellular loop region is poorly conserved, even among Pit proteins that are closely related, which provides perplexing evidence for a functional domain or residues that are required for the Pi uptake reaction.

4.7 - Other PhoU-like proteins

Many proteins were identified which contained PhoU and PhoU_div domains which were significant (E-values less than E-0-5) were not the domain with the lowest E-value. Intriguingly, PhoU_div and PhoU hits were found with Na_Pi_cotrans domains as

the best hit, a domain that is primarily found in eukaryotic Na+/Pi transport proteins (Hilfiker *et al*., 1998). However, Na+/Pi cotransporter proteins are also found in lower eukaryotes and bacteria, such as *E. coli* YjbB, which was suggested to be an exporter of Pi, as YjbB was found to reduce the effects of increased polyP accumulation in a *phoU* mutant without altering the constitutive Pho induction (Motomura *et al.,* 2011). Thus, the PhoU domains are believed modulate export activity of YjbB, rather than by complementing the *phoU* mutation. Also, PhoU_div hits were found with MFS_1 as a best hit (E-value < E-17), which is a domain found in proteins of the major facilitator superfamily transporter, which also includes Pit proteins. This supports our hypothesized association between Pap and Pit, as it suggests PhoU_div domains are involved in proteins that conduct transport of solutes into or out of the cell.

As PhoU and PhoU_div domains are found on many transport proteins, it is likely that their function is somewhat like that of Pap and PhoU in modulating the transport of the cognate transporter protein in some way. It is may be that the transport proteins are then modulated in response to sensing some intermediate molecule, whose production or catabolism is downstream of the cognate transport system.

A similar sensor-transporter model has been proposed for *Saccharomyces cerevisiae*, where the SPX domain of low-affinity Pi transporters Pho87 and Pho90 are involved in inactivation of the respective transporters by endocytosis (Ghillerbert *et al.,* 2011; Secco *et al.,* 2012). The SPX domain is found within the first 150-200 aa and has been found to interact with another protein of unknown function, Spl2 (Hurlimann *et al.,* 2009). Spl2 then acts as an adaptor protein involved in mediating interactions between the

Pho87/Pho90 transporters and coat proteins. SPX is believed to sense environmental Pi concentrations, not by binding Pi directly but through the intermediary inositol-phosphates (like inositol triphosphate, pyrophosphate or hexaphosphate), which are known metabolic signal molecules in eukaryotes (Azevedo and Saiardi, 2017).

SPX domains are typically found in large proteins with multiple protein domains, including some that also have MFS domains (Du *et al*., 2017). However, SPX domains are also found as independent proteins – for example in the SPX1 protein of *Arabidopsis thaliana* (Puga *et al*., 2014). This may be similar to PhoU and Pap domains that are found as individual proteins that modulate a cognate transporter. Crystal structures for five proteins that contain SPX domains have been determined (PBD: 5IIG, 5LNC, 5IJH, 5IJJ, 5IJP). Interestingly these structures reveal that the SPX domain is made-up of three large α-helices and three small α-helices that form a parallel helical bundle, reminiscent to that of PhoU and Pap (Wild *et al.*, 2016).

Because the intracellular loop appears to be a region of interest in MFS and Pit family proteins, it may be that this loop region confers a similar function as PhoU-like domains. In a structural prediction of the EcPitA loop region, the putative structure consisted of 3 parallel α-helices separated by short loop regions, however with poor accuracy as there were no similar proteins in the PDB (Estimated RMSD 9.9±4.6Å) (Appendix Figure 11). Intriguingly, the SPX domain of *S. cerevisiae* polyP polymerase Vtc4 was identified as the PDB-hit with the best structural alignment (TM-Score: 0.664, RMSD 3.42Å). Thus, this apparent coincidental similarity may warrant purification and crystallization of the intracellular loop region of Pit proteins.

4.8 - A model of PhoU-like proteins and the cognate transporter proteins

The evidence for enzyme modulation in yeast via the SPX domain is promising as a model for PhoU-like proteins. Perhaps, like the SPX domain, PhoU and PhoU-like domains are involved in modulation of Pi uptake via an intra-protein interaction (such as the case with Na+/Pi-cotransporters with the PhoU domains) or an interprotein interaction with the cognate transporter (whether PstSCAB or Pit). diCenzo *et al*. (2017) proposed that PhoU is involved in sensing of intracellular Pi concentrations, permitting rapid deactivation of PstSCAB in high Pi environments to prevent excess Pi accumulation which is lethal to the cell.

As a hypothesis, PhoU and PhoU-like proteins may sense cytoplasmic Pi through a small intermediate molecule, such as a phosphate metabolite molecule containing phosphate, which would be indicative of high intracellular Pi. The interaction with the small molecule would then results in a conformational change which would alters protein-protein interactions with the cognate transport system and alter the rate or affinity of the transport reaction.

The conserved nature of the PhoU domains suggests proteins with this domain share Pi sensing capability, especially to a signal molecule that would indicate high intracellular Pi concentrations, which appears to be the response observed for Pap and PhoU. The difference between Pap and PhoU thus lies solely within the interaction with its cognate transporter and not the sensing and signaling interactions. This heterologous interaction could be mediated by altering dimerization of the transporter proteins to either

promote (Pit or Na+/Pi-cotransporter dimers) or disrupt (PstB-PstB dimer) interactions, which are imperative for the function of the uptake system. (Figure 4-1).

This hypothesis provides a model of rapid modulation of activity, rather than the expression (transcription/translation) of the Pi transport proteins in response to altered cytoplasmic Pi concentration. Further experimentation, particularly into determining the interaction between PhoU-like protein and cognate transporter, as well as whether the PhoU-domain binds to a specific ligand, is necessary. The uptake modulation due to protein-protein interaction between the PhoU-like protein and cognate transporter would also need to be explored, which could be determined by crystallography of the protein-protein interaction at open-closed intermediate states. Determining this molecular mechanism is imperative for understanding the precise details of Pi homeostasis in microorganisms particularly for the significant applications of phosphate metabolism or catabolism in biotechnology.

4.9 – Relevance and Biotechnological Applications

Several proteomes were found to have multiple copies of the Pap-Pit operon, in addition to other Pit orthologues and multiple copies of the PstSCAB-PhoU genes. From our analysis, three putative *pap*-pit orthologues were identified from metagenome-reconstituted genome of *Candidatus Accumulibacter phosphatis* (Martin *et al.*, 2006). *A. phosphatis* is a Pi-accumulating organism implicated as a major organism involved in enhanced biological phosphorus removal (EBPR) in wastewater treatment facilities (He *et al*., 2010). Whether any of these orthologues play a critical role in Pi accumulation has

yet to be determined, but may contribute to the effectiveness of Pi accumulation in EBPR. In addition, PhoU-like proteins are of interest because mutations in *phoU* result in constitutive Pho regulon activation and polyphosphate accumulation in many bacterial species. Thus, experimentation in Pap proteins could be translationally applied to PhoU proteins. By fully understanding the regulation as well as modulation of Pi uptake, assimilation and accumulation, these organisms or synthetic organisms could be used to produce polyphosphates which have many biotechnological applications, such as use as agricultural fertilizers.

Works Cited

Azevedo, C., and Saiardi, A. "Eukaryotic phosphate homeostasis: The inositol pyrophosphate perspective." *Trends in biochemical sciences 42(3)* (2017): 219-231.

Baneyx, F., and Mujacic, M. "Recombinant protein folding and misfolding in *Escherichia coli*." *Nature biotechnology 22(11)* (2004): 1399.

Bardin, S., Dan, S., Osteras, M., Finan, T.M. "A Phosphate Transport System is Required for Symbiotic Nitrogen Fixation by *Rhizobium meliloti*." *Journal of Bacteriology 178(15)* (1996): 4540-547.

Bardin, S.D., and Finan, T.M. "Regulation of phosphate assimilation in *Rhizobium (Sinorhizobium) meliloti*." *Genetics 148(4)* (1998): 1689-1700.

Beard, S.J., Hashim,R., Wu, G., Binet, M.R.B., Hughes, M.N., Pooles, R.K. "Evidence for the tranport of zinc (II) ions via the Pit inorganic phosphate transport system in *Escherichia coli*." *FEMS Microbiology Letters 184* (2000): 231-235.

Bøttger, P., and Pedersen, L. "Mapping of the minimal inorganic phosphate transporting unit of human PiT2 suggests a structure universal to PiT-related proteins from all kingdoms of life." *BMC biochemistry 12(1)* (2011): 21.

Bøttger, P., and Pedersen, L. "Two highly conserved glutamate residues critical for type III sodium-dependent phosphate transport revealed by uncoupling transport function from retroviral receptor function." *Journal of Biological Chemistry 277(45)* (2002): 4.

Chan, F.Y., and Torriani, A. "PstB protein of the phosphate-specific transport system of *Escherichia coli* is an ATPase." *Journal of Bacteriology 178(13)* (1996): 3974-3977.

Cox, G.B., Webb, D., Rosenberg, H. "Specific amino acid residues in both the PstB and PstC proteins are required for phosphate transport by the *Escherichia coli* Pst system." *Journal of bacteriology 171(3)* (1989): 1531-1534.

Crépin, S., Chekabab, S.M., Le Bihan, G., Bertrand, N., Dozois, C.M., Harel, J. "The Pho Regulon and the pathogenesis of *Escherichia coli*." *Veteranary Microbiology 153(1)* (2011): 82-88.

Dalbey, R.E., Wang, P., and Kuhn, A. "Assembly of bacterial inner membrane proteins." *Annual review of biochemistry 80* (2011): 161-187.

de Almeida, L.G., Ortiz, J.H., Schneider, R.P., and Spira, B. "*phoU* inactivation in *Pseudomonas aeruginosa* enhances accumulation of ppGpp and polyphosphate." *Applied and environmental microbiology 81(9)* (2015): 3006-3015.

dicenzo, G.C., Sharthiya, H., Nanda, A., Zamani, M., & Finan, T.M. "PhoU allows rapid adaptation to high phosphate concentrations by modulating PstSCAB transport rate in *Sinorhizobium meliloti*." *Journal of Bacteriology* (2017): JB-00143.

Du, H., Yang, C., Ding, G., Shi, L., Xu, F. "Genome-wide identification and characterization of SPX domain-containing members and their responses to phosphate deficiency in *Brassica napus*." *Front Plant Sci 8* (2017): 35.

Elvin, C.M., Dixon, N.E., Rosenberg, H. "Molecular cloning of the phosphate (inorganic) transport (pit) gene in *Escherichia* K12." *Molecular Genetics and Genomics 204* (1986): 477-484.

Gardner, S.G., Johns, K.D., Tanner, R., and McCleary, W.R. "The PhoU protein from *Escherichia coli* interacts with PhoR, PstB, and metals to form a phosphate-signaling complex at the membrane." *Journal of bacteriology 196(9)* (2014): 1741-1752.

Gardner, S.G., Miller, J.B., Dean, T., Robinson, T., Erickson, M., Ridge, P.G., & McCleary, W.R. "Genetic analysis, structural modeling, and direct coupling analysis suggest a mechanism for phosphate signaling in *Escherichia coli*." *BMC genetics 16(2)* (2015): S2.

Gebhard, S., Tran, S.L., and Cook, G.M. "The Phn system of *Mycobacterium smegmatis*: a second high-affinity ABC-transporter for phosphate." *Microbiology 152(11)* (2006): 3453-3465.

Geiger, O., Röhrs, V., Weissenmayer, B., Finan, T.M., & Thomas-Oates, J.E. "The regulator gene *phoB* mediates phosphate stress-controlled synthesis of the membrane lipid diacylglyceryl-N, N, N-trimethylhomoserine in *Rhizobium (Sinorhizobium) meliloti*." *Molecular microbiology 32(1)* (1999): 63-73.

Ghillebert, R., Swinnen, E., De Snijder, P., Smets, B., and Winderickx, J. "Differential roles for the low-affinity phosphate transporters Pho87 and Pho90 in *Saccharomyces cerevisiae*." *Biochemical Journal 434(2)* (2011): 243-251.

He, S., Kunin, V., Haynes, M., Martin, H.G., Ivanova, N., Rohwer, F., McMahon, K.D. "Metatranscriptomic array analysis of '*Candidatus Accumulibacter phosphatis*'-enriched enhanced biological phosphorus removal sludge." *Environmental microbiology 12(5)* (2010): 1205-1217.

Hilfiker, H., Hattenhauer, O., Traebert, M., Forster, I., Murer, H., and Biber, J. "Characterization of a murine type II sodium-phosphate cotransporter expressed in mammalian small intestine." *Proceedings of the National Academy of Sciences, 95(24)* (1998): 14564-14569.

Höglund, P.J., Nordström, K.J., Schiöth, H.B., and Fredriksson, R. "The solute carrier families have a remarkably long evolutionary history with the majority of the human families present before divergence of Bilaterian species." *Molecular biology and evolution 28(4)* (2010): 1531-1541.

Hsieh, Y.Y., Wanner, B.L. "Global regulation by the seven-component Pi signaling system." *Current Opinion in Microbiology 13(2)* (2010): 198-203.

Hürlimann, H.C., Pinson, B., Stadler-Waibel, M., Zeeman, S.C., & Freimoser, F.M. "The SPX domain of the yeast low-affinity phosphate transporter Pho90 regulates transport activity." *EMBO reports 10(9)* (2009): 1003-1008.

"Inorganic phosphate and sulfate transport in *S. cerevisiae*." Samyn, D.R., & Persson, B.L. *Yeast Membrane Transport* . Springer International Publishing., 2016. 253-269.

Islam, S.T., Taylor, V.L., Qi, M., and Lam, J.S. " Membrane topology mapping of the O-antigen flippase (Wzx), polymerase (Wzy), and ligase (WaaL) from *Pseudomonas aeruginosa* PAO1 reveals novel domain architectures." *MBio 1(3)* (2010): e00189-10.

Jeong, J.Y., Yim, H.S., Ryu, J.Y., Lee, H.S., Lee, J.H., Seen, D.S., and Kang, S.G. "One-step sequence-and ligation-independent cloning as a rapid and versatile cloning method for functional genomics studies." *Applied and environmental microbiology 78(15)* (2012): 5440-5443.

Jono, S., McKee, M.D., Murry, C.E., Shioi, A., Nishizawa, Y., Mori, K., and Giachelli, C.M. "Phosphate regulation of vascular smooth muscle cell calcification." *Circulation research 87(7)* (2000): e10-e17.

Karimova, G., Pidoux, J., Ullmann, A., and Ladant, D. "A bacterial two-hybrid system based on a reconstituted signal transduction pathway." *Proceedings of the National Academy of Sciences 95(10)* (1998): 5752-5756.

Liu, J., Lou, Y., Yokota, H., Adams, P.D., Kim, R., and Kim, S.H. "Crystal Structure of a PhoU Protein Homologue: A new class of metaloprotein containing multinuclear iron clusters." *Journal of Biological Chemistry 280(16)* (2005): 15960-15966.

Lubin, E.A., Henry, J.T., Fiebig, A., Crosson, S., Laub, M.T. "Identification of the PhoB regulon and role of PhoU in the phosphate starvation response of *Caulobacter crescentus*." *Journal of bacteriology 198(1)* (2016): 187-200.

MacArthur, M.W. and Thornton J.M. "Influence of Proline Residues on Protein Conformation." *Journal of molecular biology 218* (1991): 397-412.

Makino, K., Amemura, M., Kawamoto, T., Kimura, S., Shinagawa, H., Nakata, A., Suzuki, M. "DNA Binding of PhoB and its Interaction with RNA Polymerase." *Journal of Molecular Biology 259(1)* (1996): 15-26.

Makino, K., Shinagawa, H., Amemura, M., Kawamoto, T., Yamada, M., Nakata, A. "Signal Transduction in the Phosphate Regulon of *Escherichia coli* Involves Phosphotransfer Between PhoR and PhoB Proteins." *Journal of Molecular Biology 210* (1989): 551-559.

Martin, H.G., Ivanova, N., Kunin, V., Warnecke, F., Barry, K.W., McHardy A.C., Hugenholtz, P. "Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities." *Nature Biotechnology 24(1)* (2006): 1263-1269. .

Medveczky, N. and Rosenberg, H. "Phosphate tranport in *Escherichia coli*." *Biochimica et Biophysica Acta* (1971): 494-506.

Metcalf, W.W., and Wanner, B.L. "Evidence for a fourteen-gene, *phnC* to *phnP* locus for phosphonate metabolism in *Escherichia coli*." *Gene 129(1)* (1993): 27-32.

Milunovic, B., Morton, R.A., and Finan, T.M. "Cell growth inhibition upon deletion of four toxin-antitoxin loci from the megaplasmids of *Sinorhizobium meliloti*." *Journal of bacteriology 196(4)* (2014): 811-824.

Morohoshi, T., Maruo, T., Shirai, Y., Kato, J., Ikeda, T., Takiguchi, N., and Kuroda, A. "Accumulation of inorganic polyphosphate in *phoU* mutants of *Escherichia coli* and *Synechocystis sp.* strain PCC6803." *Applied and environmental microbiology 68(8)* (2002): 4107-4110.

Motomura, K., Hirota, R., Ohnaka, N., Okada, M., Ikeda, T., Morohoshi, T., and Kuroda, A. "Overproduction of YjbB reduces the level of polyphosphate in *Escherichia coli*: a hypothetical role of YjbB in phosphate export and polyphosphate accumulation." *FEMS microbiology letters 320(1)* (2011).

Oganesyan, V., Oganesyan, N., Adams, P.D., Jancarik, J., Yokota, H.A., Kim, R., and Kim, S.H. "Crystal structure of the "PhoU-like" phosphate uptake regulator from *Aquifex aeolicus*." *Journal of bacteriology 187(12)* (2005): 4238-4244.

Okuda, S., Kawashima, S., Kobayashi, K., Ogasawara, N., Kanehisa, M., Goto, S. "Characterization of relationships between transcriptional units and operon structures in *Bacillus subtilis* and *Escherichia coli*." *BMC Genomics 8* (2007): 48.

Oresnik, I.J., Charles, T.C., Finan, T.M. "Second Site Mutations Specifically Suppress the Fix- Phenotype of *Rhizobium meliloti ndvF* Mutations on Alfalfa: Identification of a Conditional ndvF-Dependent Mucoid Colony Phenotype." *Genetics* (1994): 1233-1243.

Privé, G.G. "Detergents for the stabilization and crystallization of membrane proteins." *Methods 41(4)* (2007): 388-397.

Puga, M.I., Mateos, I., Charukesi, R., Wang, Z., Franco-Zorilla, J.M., Paz-Ares, J. "SPX1 is a phosphate-dependent inhibitor of PHOSPHATE STARVATION RESPONSE 1 in Arabidopsis." *Proceedings of the National Academy of Science USA 111(41)* (2014): 14947-14952.

Quandt, J., and Hynes, M.F. "Versatile suicide vectors which allow direct selection for gene replacement in gram-negative bacteria." *Gene 127(1)* (1993): 15-21.

Reddy, V.S., Shlykov, M.A., Castillo, R., Sun, E.I., and Saier, M.H. " The major facilitator superfamily (MFS) revisited." *The FEBS journal 279(11)* (2012): 2022-2035.

Rice, C.D., Pollard, J.E., Lewis, Z.T., McCleary, W.R. "Employment of a promoter-swapping technique shows that PhoU modulates the activity of the PstSCAB2 ABC transporter in *Escherichia coli*." *Applied and environmental microbiology 75(3)* (2009): 573-582.

Rosenberg, H., Gerdes, R.G., and Chegwidden, K. "Two systems for the uptake of phosphate in *Escherichia coli*." *Journal of Bacteriology 131(2)* (1977): 505-511.

Rosenberg, H., Gerdes, R.G., Harold, F.M. "Energy coupling to the transport of inorganic phosphate in *Escherichia coli* K12." *Biochemical Journal 178* (1979): 133-137.

Russell, L.M. and Rosenberg, H. "The nature of the link between potassium transport and phosphate transport in *Escherichia coli*." *Biochemical Journal 188* (1980): 715-723.

Russell, L.M., and Rosenberg, H. "Linked transport of phosphate, potassium ions and protons in *Escherichia coli*." *Biochemical Journal 184* (1979): 13-21.

Saier, M.H., Eng, B.H., Fard, S., Garg, J., Haggerty, D.A., Hutchinson, W.J., and Omar, A.M. "Phylogenetic characterization of novel transport protein families revealed by genome analyses." *Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes 1422(1)* (1999): 1-56.

Salaün, C., Gyan, E., Rodrigues, P., and Heard, J.M. "Pit2 assemblies at the cell surface are modulated by extracellular inorganic phosphate concentration." *Journal of virology 76(9)* (2002): 4304-4311.

Salaün, C., Rodrigues, P., and Heard, J.M. "Transmembrane topology of PiT-2, a phosphate transporter-retrovirus receptor." *Journal of virology 75(12)* (2001): 5584-5592.

Samyn, D.R., Ruiz-Pávon, L., Andersson, M.R., Popova, Y., Thevelein, J.M., and Persson, B.L. " Mutational analysis of putative phosphate-and proton-binding sites in the *Saccharomyces cerevisiae* Pho84 phosphate: H+ transceptor and its effect on signalling to the PKA and PHO pathways." *Biochemical Journal 445(3)* (2012): 413-422.

Secco, D., Wang, C., Shou, H., and Whelan, J. "Phosphate homeostasis in the yeast *Saccharomyces cerevisiae*, the key role of the SPX domain-containing proteins." *FEBS letters 586(4)* (2012): 289-295.

Sengottaiyan, P., Ruiz-Pavón, L., and Persson, B.L. "Functional expression, purification and reconstitution of the recombinant phosphate transporter Pho89 of *Saccharomyces cerevisiae*." *The FEBS journal 280(3)* (2013): 965-975.

Shieh, Y.W., Minguez, P., Bork, P., Auburger, J.J., Guilbride, D.L., Kramer, G., and Bukau, B. "Operon structure and cotranslational subunit association direct protein assembly in bacteria." *Science 350(6261)* (2015): 678-680.

Steed, P.M., and Wanner, B.L. "Use of the rep technique for allele replacement to construct mutants with deletions of the *pstSCAB-phoU* operon: evidence of a new role for the PhoU protein in the phosphate regulon." *Journal of bacteriology 175(21)* (1993): 6.

Surin, B.P., Dixon, N.E., and Rosenberg, H. "Purification of the phoU protein, a negative regulator of the *pho* regulon of *Escherichia coli* K-12." *Journal of bacteriology 168(2)* (1986): 631-635.

van de Guchte, M., Kok, J., and Venema, G. "Distance-dependent translational coupling and interference in *Lactococcus lactis*." *Molecular Genetics and Genomes 227* (1991): 65-71. .

van Veen, H.W. "Phosphate transport in prokaryotes: molecules, mediators and mechanisms." *Antonie Van Leeuwenhoek 72(4)* (1997): 299-315.

Versaw, W.K., and Metzenberg, R.L. "Repressible cation-phosphate symporters in Neurospora crassa." *Proceedings of the National Academy of Sciences 92(9)* (1995): 3884-3887.

Voegele, R.T., Bardin, S., and Finan, T.M. "Characterization of the *Rhizobium (Sinorhizobium) meliloti* high-and low-affinity phosphate uptake systems." *Journal of bacteriology 179(23)* (1997): 7226-7232.

Wang, B., Rensing, C., Pierson III, L.S., Zhao, H., Kennedy, C. "Translational coupling of *nasST* expression in *Azotobacter vinelandii* prevents overexpression of the nasT gene." *FEMS Microbiology Letters 361* (2014): 123-130.

Weinglass, A.B., and Kaback, H.R. "The central cytoplasmic loop of the major facilitator superfamily of transport proteins governs efficient membrane insertion." *Proceedings of the National Academy of Sciences 97(16)* (2000): 8938-8943.

Wells, J.N., Bergendahl, L.T., and Marsh, J.A. "Operon gene order is optimized for ordered protein complex assembly." *Cell reports 14(4)* (2016): 679-685.

Wheeler, T.J., Clements, J., and Finn, R.D. "Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models." *BMC Bioinformatics 15* (2014): 7.

White, C.E., Gavina, J.M., Morton, R., Britz-McKibbin, P., Finan, T.M. "Control of hydroxyproline catabolism in *Sinorhizobium meliloti*." *Mol Microbiol 85(6)* (2012): 1133-1147. .

Wild, R., Gerasimaite, R., Jung, J.Y., Truffault, V., Pavlovic, I., Schmidt, A., and Mayer, A. (2016). "Control of eukaryotic phosphate homeostasis by inositol polyphosphate sensor domains." *Science aad9858.* (2016).

Williams, K.P., Gillespie, J.J., Sobral, B.W.S., Nordberg, E.K.. Snyder, E.E.. Shallom, J.M., Dickerman, A.W. "Phylogeny of Gammaproteobacteria." *Journal of bacteriology 192(9)* (2010): 2305-2314.

Willsky, G.R., and Malamy, M.H. "Characterization of two genertically separable inorganic phosphate transport systems in *Escherichia coli*." *Journal of Bacteriology 144(1)* (1980): 356-365.

Wykoff, D.D., and O'Shea, E.K. ". Phosphate transport and sensing in *Saccharomyces cerevisiae*." *Genetics 159(4)* (2001): 1491-1499.

Xu, L., Li, Y., Haworth, I.S., and Davies, D.L. "Functional role of the intracellular loop linking transmembrane domains 6 and 7 of the human dipeptide transporter hPEPT1." *The Journal of membrane biology 238(1-3)* (2010): 43-49.

Yakovleva, G.M., Kim, S.K., and Wanner, B.L. "Phosphate-independent expression of the carbon-phosphorus lyase activity of *Escherichia coli*." *Applied microbiology and biotechnology 49(5)* (1998): 573-578.

Yan, N. "Structural Biology of the Major Facilitator Superfamily Transporters." *Annual Review of Biophysics 44* (2015): 257-283.

Yuan, Z.-C., Zaheer, R., Finan, T.M. "Regulation and Properties of PstSCAB, a High-Affinity, High-Velocity Phosphate Transport System of *Sinorhizobium meliloti*." *Journal of Bacteriology 188(3)* (2006): 1089-1102.

Yuan, Z.-C., Zaheer, R., Morton, R., Finan, T.M. "Genome prediction of PhoB regulated promoters in *Sinorhizobium meliloti* and twelve proteobacteria." *Nucleic Acid Research 34(9)* (2006): 2686-2697.

Zaheer, R., Morton, R., Proudfoot, M., Yakunin, A., and Finan, T.M. "Genetic and biochemical properties of an alkaline phosphatase PhoX family protein found in many bacteria." *Environmental microbiology 11(6)* (2009): 1572-1587.

Zvyagilskaya, R.A., Lundh, F., Samyn, D., Pattison-Granberg, J., Mouillon, J.M., Popova, Y., andPersson, B.L. "Characterization of the Pho89 phosphate transporter by functional hyperexpression in *Saccharomyces cerevisiae*." *FEMS yeast research 8(5)* (2008): 685-696.

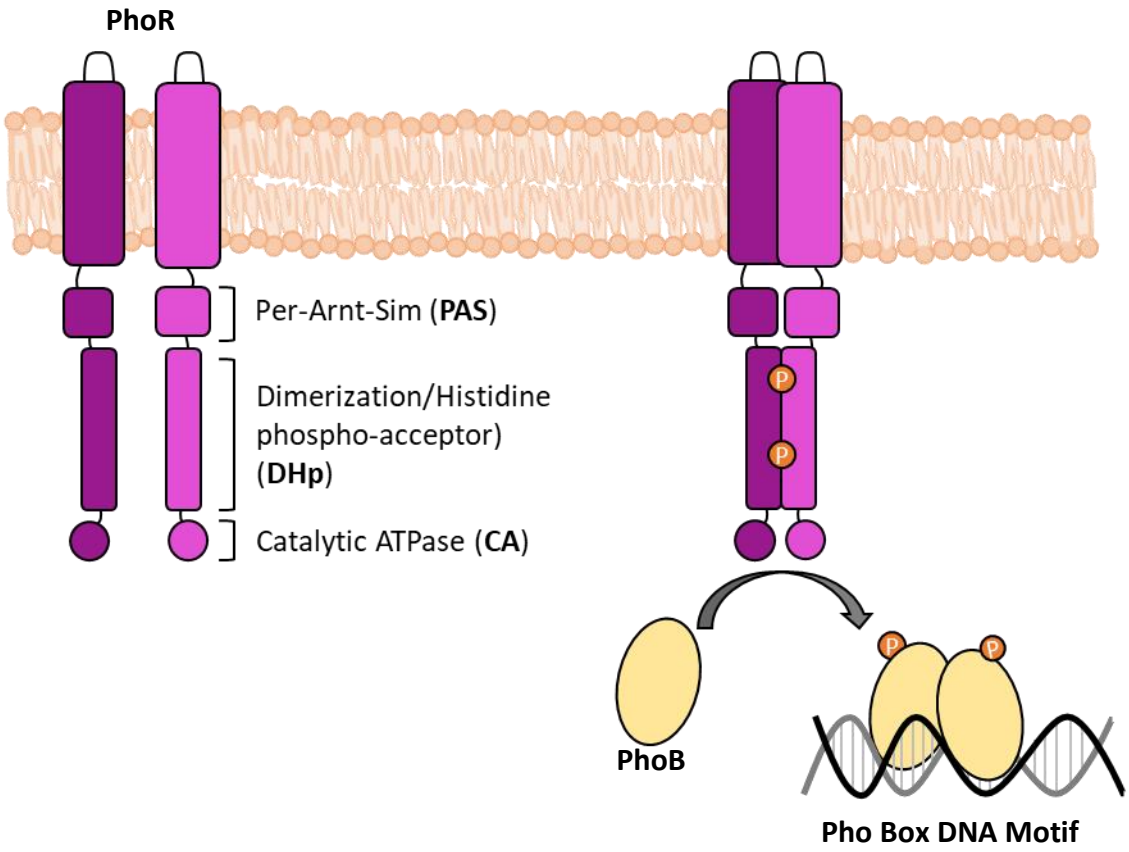


Figure 1-1. Diagram of the PhoR-PhoB two component system. PhoR (purple) integral membrane proteins contain a short periplasmic region separated by two transmembrane helices and large C-terminal catalytic tail. The tail region contains a PAS domain, possible signal sensor of Pi (orange circles), dimerization/Histidine phospho-acceptor domain, where the His-residues are autophosphorylated and/or dephosphorylated by the catalytic ATPase domain. PhoR senses low intracellular phosphate concentrations and upon transfer of a phosphoryl group to PhoB (yellow), which dimerizes and binds to Pho-box motifs of specific genes.

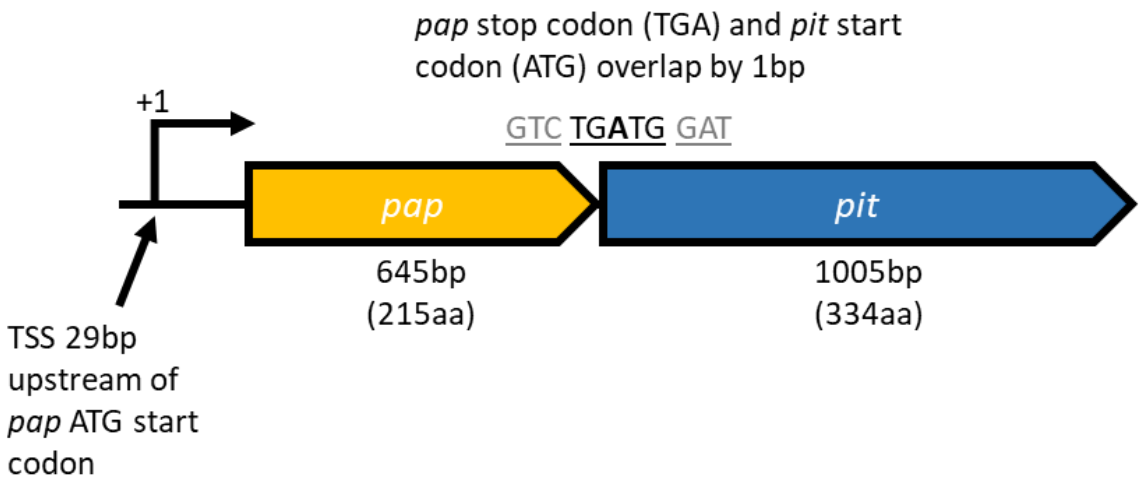


Figure 1-2. Diagram of the *S. meliloti pap-pit* operon. *Sinorhizobium pap-pit* is encoded as an operon with coding regions that overlap by a single bp. *pap-pit* encodes proteins of 215 and 334 a.a., respectively. The *S. meliloti pap-pit* operon is negatively regulated by PhoB under low Pi conditions. The transcriptional start site of the operon is situated 29 bp upstream of the *pap* ATG, which was previously determined by RNAseq of wildtype *S. meliloti* RmP110 in MOPS-P2 (high Pi conditions) (Milunovic *et al*., 2011).

Figure 1-3. Diagram of Pit proteins and PHO4 domains. Pit proteins have two PHO4 domains, believed to have occurred due to a tandem duplication event.

Table 2-1. Table of Strains and plasmids used in this work.

Sinorhizobium meliloti

Strain	Description	Source or Reference
Rm1021	*S. meliloti* SU47-derived Sm resistant isolate with *pstC* frameshift mutation	
RmP110	Rm1021 with *pstC* mutation corrected	Yuan *et al.,* 2006
RmP1628	RmP110 (pTH2340 - Δ*pap*)	Zaheer, Unpublished
RmP1629	RmP110 (pTh2341 - Δ*pit*)	Zaheer, Unpublished
RmP3151	RmP110 (Δ*pap-pit*)	Zaheer and Zhao, Unpublished
RmP3421	P_{lac}::*pstSCAB-phoUB*, P_{lac}::*phoR, lacIq, phoC::*ΩSp^R	diCenzo *et al.,* 2017
RmG430	Rm1021 *phe232*::Tn5 (Gm^R)	Oresnik *et al.,*
RmP3923	RmP3421 (φRmG430 *phe232::*Gm^R)	This work
RmP3924	RmP3923 (φRmP1628 Δ*pap*)	This work
RmP3925	RmP3923 (φRmP1629 Δ*pit*)	This work
RmP3926	RmP3923 (φRmP3151 Δ*pap-pit*)	This work

Escherichia coli

Strain	Description	Source or Reference
DH5α	*endA1 hsdR17 supE44 thi-1 recA1 gyrA96 relA1* Δ(*argF-lacZYA*)U169 φ80d*lacZ*ΔM15	
MG1655	Wildtype *strain* F^- *arcA-1655 fnr-1655*)	
MT2006	MG1655 Δ*pitA::frt* Δ*pitB::frt* Δ*phnC::frt* Δ(*pstSCAB-phoU*)::Km^R	Motomura *et al.* (2011)
BTH101	F^- *cya-99 araD139 galE15 galK16 rpsL1 hsdR2 mcrA1 mcrB1*	Euromedex

Plasmids

Plasmid	Description	Source or Reference
pJQ200	p15A origin vector derived from pACYCY184-Gm^R with *Bacillus subtilis sacB*, pUC18 polylinker and *lacZα* fragment, Gm^R	Quandt *et al.* (1993)
pTH2340	pJQ200 Δ*pap* deletion vector, Gm^R	Zaheer, Unpublished
pTH2341	pJQ200 Δ*pit* deletion vector, Gm^R	Zaheer, Unpublished
pUCP30T	pRO1600 origin (*Pseudomonas aeruginosa*) broad host-range cloning vector, (suicide vector *in S. meliloti*) Gm^R	Schweizer and Klassen (1995)
pTH2824	pUCP30T (*Sm pap-pit* with 500bp upstream and downstream) via XbaI/HindIII, Gm^R	Zaheer and Zhao, Unpublished
pTH2897	pUCP30T (*Sm* Δ*pap-pit*::*frt* with 500bp upstream and downstream) via XbaI/HindIII, Gm^R	Zaheer and Zhao, Unpublished
pTH3153	pUCP30T (*Sm pap-pit*~His10), Gm^R	This work
pTH3161	pUCP30T (*Sm* promoter-*pit*~His10), Gm^R	This work
pTH3192	pUCP30T (*Sm promoter-pap*) via XbaI/HindIII (216907-215918), Gm^R	This work
pTH3193	pUCP30T (*Sm promoter-pap-pit*^{truncated}) via XbaI/HindIII (216907-215620), Gm^R	This work
pTH3194	pUCP30T (*Sm pap*^{truncated}*-pit*) via XbaI/HindIII (216238-214918), Gm^R	This work
pTH3195	pUCP30T (*Sm promoter-pit*) from pTH3148 via XbaI/HindIII (216907-216563 and 214918), Gm^R	This work
pPLE01	pBluescript II SK(+) with recombinant alkaline phosphatase/β-galactosidase (*phoA/lacZ*) fusion, Amp^R	Islam *et al.* (2010)
pTH3196	pUCP30T (promoter-*pit*~*phoA/lacZ*) via XbaI/HindIII, Gm^R	This work
pTH3198	pUCP30T (*pap-pit*~*phoA/lacZ*) via XbaI/HindIII, Gm^R	This work
pTH3200	pUCP30T (*pap-pit*~Rho1D4) , Gm^R	This work
pTH3201	pUCP30T (*pap-pit*~FLAG), Gm^R	This work

pTH3202	pUCP30T (*pap-pit*~StrepII), Gm^R	This work
pTH3227	pUCP30T (*pcaD* protocatechuate-driven promoter) via EcoRI/BamHI, Gm^R	This work
pMW119	pSC101 Low copy-number cloning vector, Amp^R	Watahiki and Hayashi (1997)
pTH2828	pMW119 (SmPapPit via XmaI/HindIII), Amp^R	Zaheer and Zhao, Unpublished
pTH2829	pMW119 (SmPit via XmaI/HindIII), Amp^R	Zaheer and Zhao, Unpublished
pTH2872	pMW119 (SoPapPit via XmaI/HindIII), Amp^R	Zaheer and Zhao, Unpublished
pTH2873	pMW119 (SoPit via XmaI/HindIII), Amp^R	Zaheer and Zhao, Unpublished
pTH3177	pMW119 (BtPapPit via XmaI/HindIII), Amp^R	This work
pTH3223	pMW119 (BtPit via XmaI/HindIII), Amp^R	This work
pTH3224	pMW119 (EcPitA via XmaI/HindIII), Amp^R	This work
pTH1937	p15A origin vector derived from pACYC77, FRT-recombinase sequence and *oriT* from RK2, Km^R	Milunovic *et al.* (2014)
pTH3232	pTH1937 (Sp^R gene from pGE1 via PstI), Km^R Sp^R	This work
pTH3233	pTH3232 (Plac::SoPap from pTH2872), Km^R Sp^R	This work
pTH3234	pTH3232 (Plac::BtPap from pTH3177), Km^R Sp^R	This work

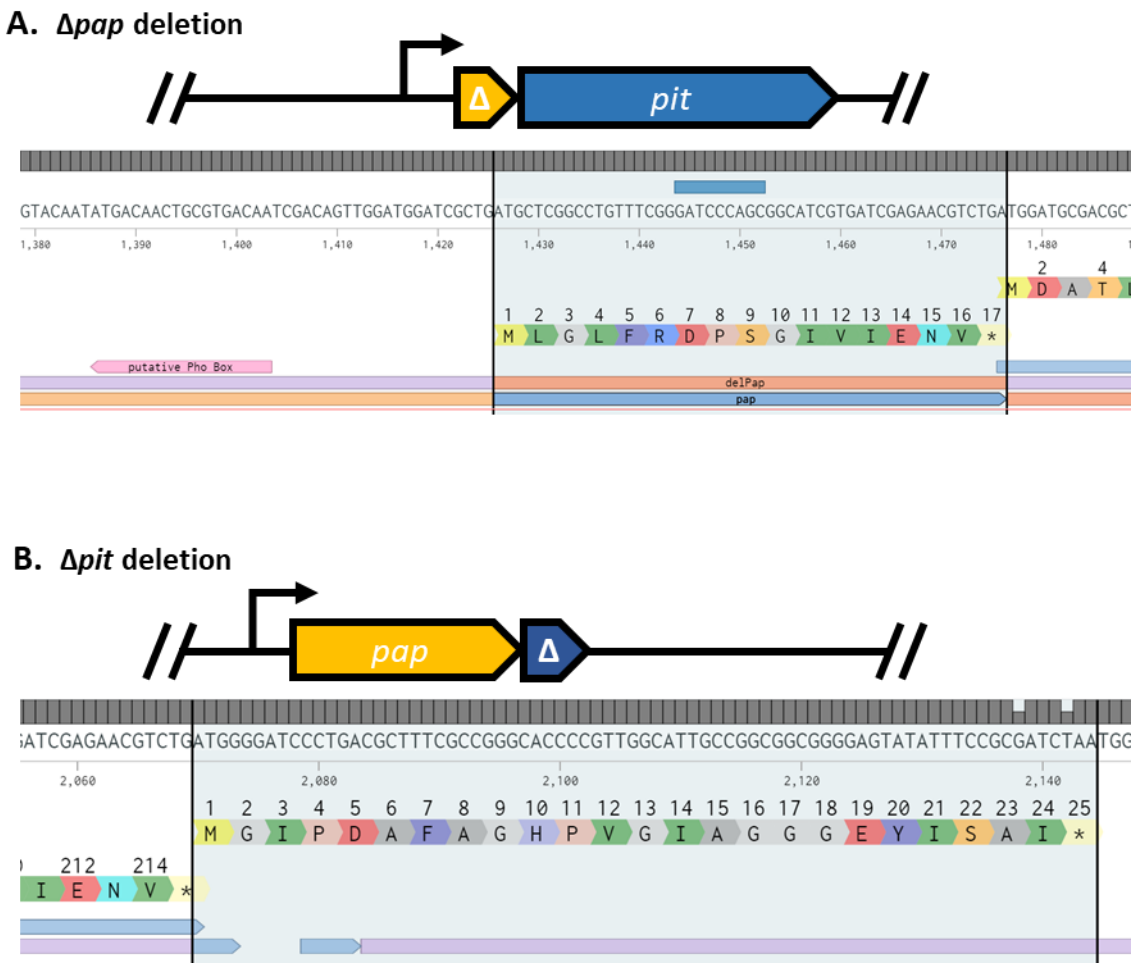


Figure 2-2. Diagram of the Δ*pap* and Δ*pit* deletions. *Sinorhizobium meliloti Δpap* and *Δpit* deletion mutations were generated via *sacB*-mediated double-crossover to counter select against single recombinants. These deletions were confirmed by Sanger sequencing as shown above.

A. *Δpap* mutation results in a truncation of SmPapK7-I206 with a BamHI restriction site, encoding a protein of only 17 a.a. The transcriptional start site of the operon is as wildtype.

B. *Δpit* mutation results in a truncation of SmPit2-334, replaced with a BamHI restriction site, encoding a protein of 25 a.a. due to a native TAA site downstream of the *pap-pit* operon.

(*** This work was done by Dr. Rahat Zaheer).

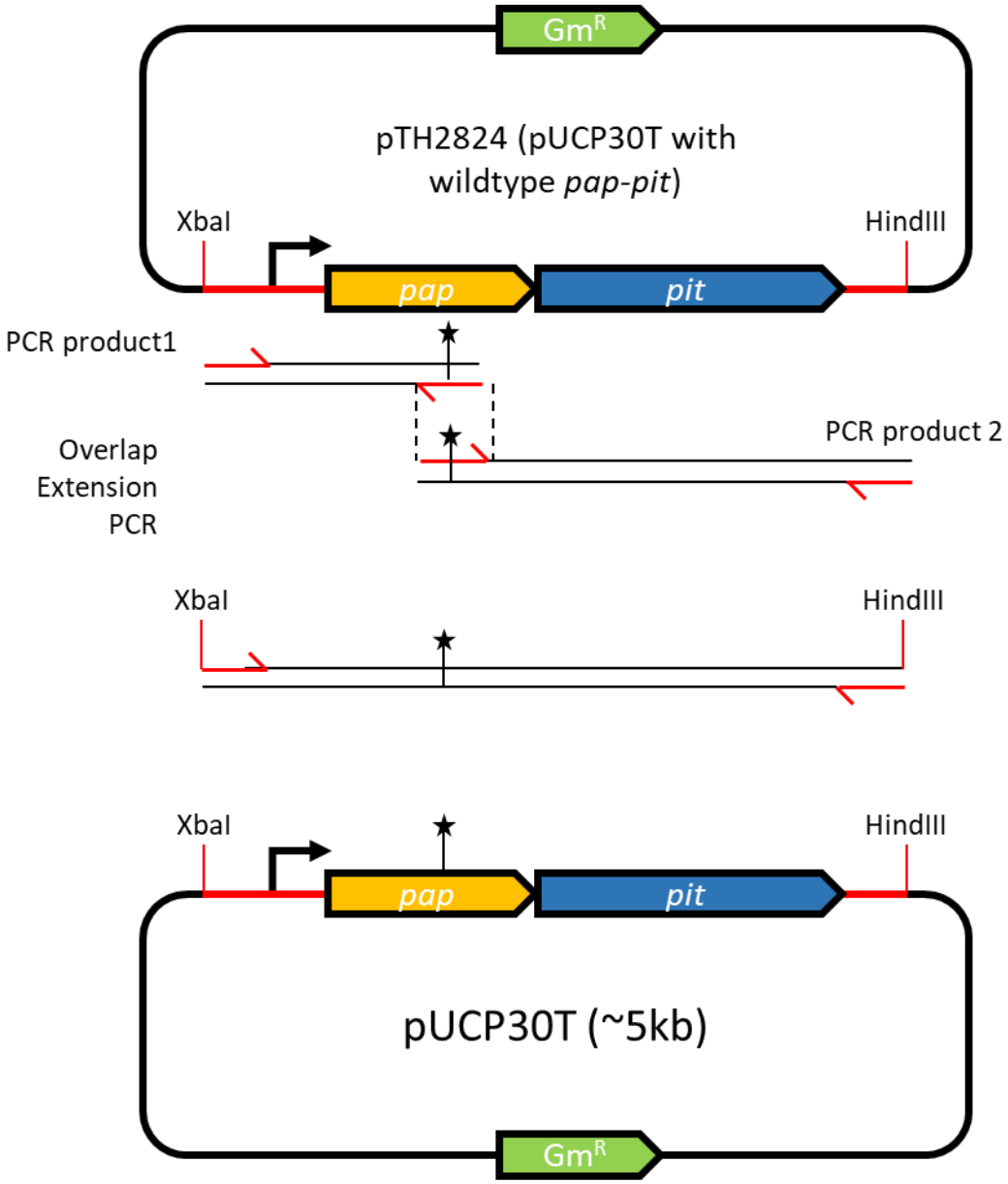


Figure 2-3. Diagram of Pap missense mutations by overlap extension PCR. Two PCR products that share a 35-40bp region of homology were amplified with primers that contained a mutation resulting in a directed missense mutation of the Pap coding sequence (star). The PCR products were then joined by overlap extension PCR (Methods 2.2.4), then subsequently cloned into pUCP30T via XbaI and HindIII restriction sites. (***This work was done by Dr. Vahid Husseininaveh).

Table 2-4. List of pUCP30T and pUC119 plasmids containing *pap-pit* with Pap missense mutations

Pap missense mutation	pUCP30T plasmid[A]	pUC119 plasmid [B]
Wildtype	pTH2824	pTH2826
E54A	pTH2094	pTH3091
D58A	pTH2095	pTH3092
T74A	pTH3040	pTH3093
P75A	pTH2906	pTH3094
R78A	pTH3043	pTH3095
D90A	pTH2907	pTH3096
D93A	pTH3039	pTH3097
D160A	pTH3035	pTH3098
E193A	pTH3036	pTH3099
D197A	pTH3037	pTH3100
D201A	pTH3038	pTH3101
D90A,H96Y	pTH2914	pTH3102
P75mis	pTH3044	pTH3103

Table 2-4. Table of Pap missense mutations which were generated and the plasmids. Pap missense mutations were cloned into pUCP30T by overlap extension PCR. These mutations were subsequently subcloned into pUC119 for expression in *E. coli*. **A.** pUCP30T plasmids contain the *pap-pit* operon with 500bp upstream and downstream, and differ only the by the point mutation. **B.** pUC119 plasmids contain lack 500bp upstream of the *pap* gene and include only 500bp downstream. The *pap-pit* operons are thus driven by the *lac* promoter from pUC119.

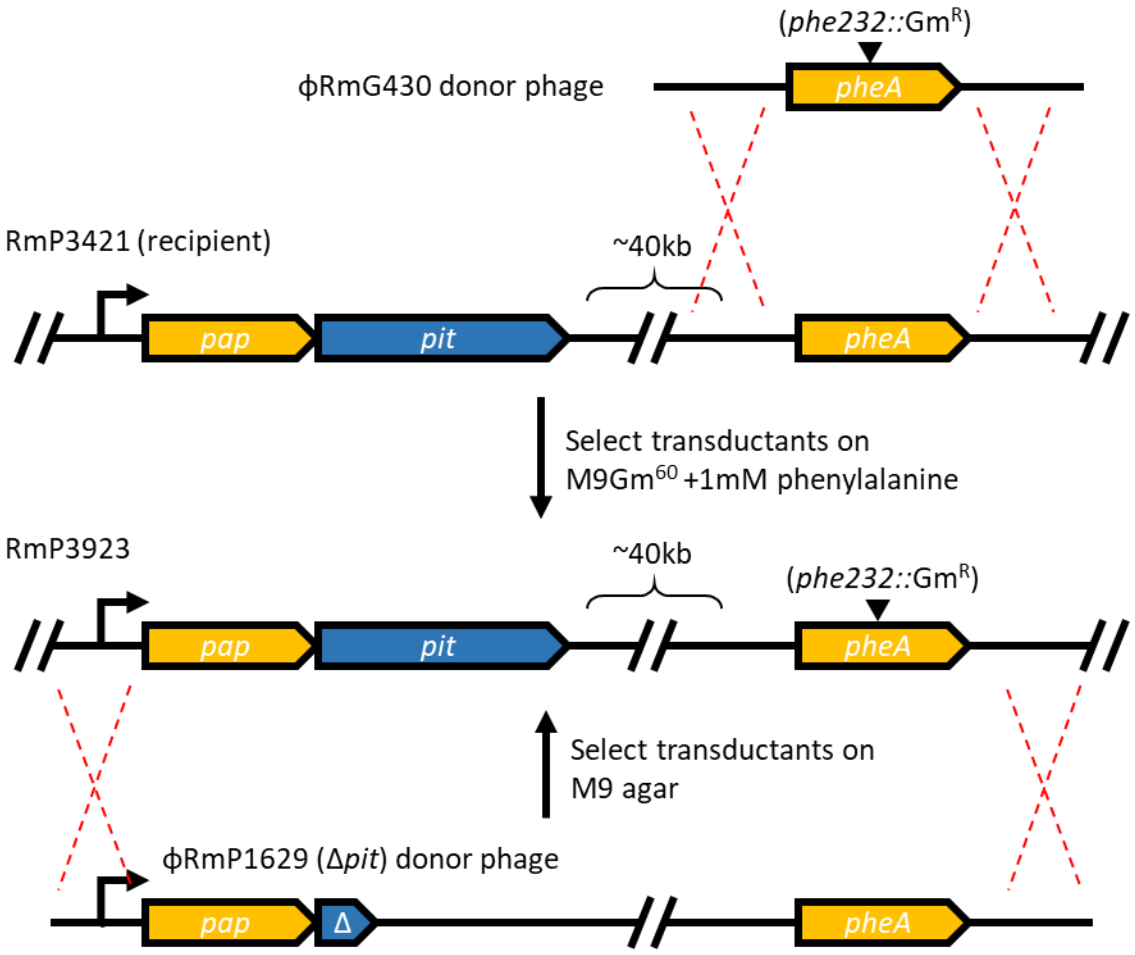


Figure 2-5. Diagram of transduction of the *pap-pit* deletions in the conditionally Pi-uptake deficient strain of *S. meliloti*. φRmG430 (*phe232::Gm*^{R}) phenylalanine auxotroph marker (probably *pheA* which is 40kb downstream of *pap-pit*) was transduced into RmP3421 to generated RmP3923 by selecting on M9 Gm supplemented with 1mM phenylalanine. Next, the unmarked Δ*pap*, Δ*pit*, and Δ*pap-pit* deletions were transduced into RmP3923 selecting for phenylalanine prototrophs on M9. Subsequent transconjugants were screened for Gm-sensitivity.

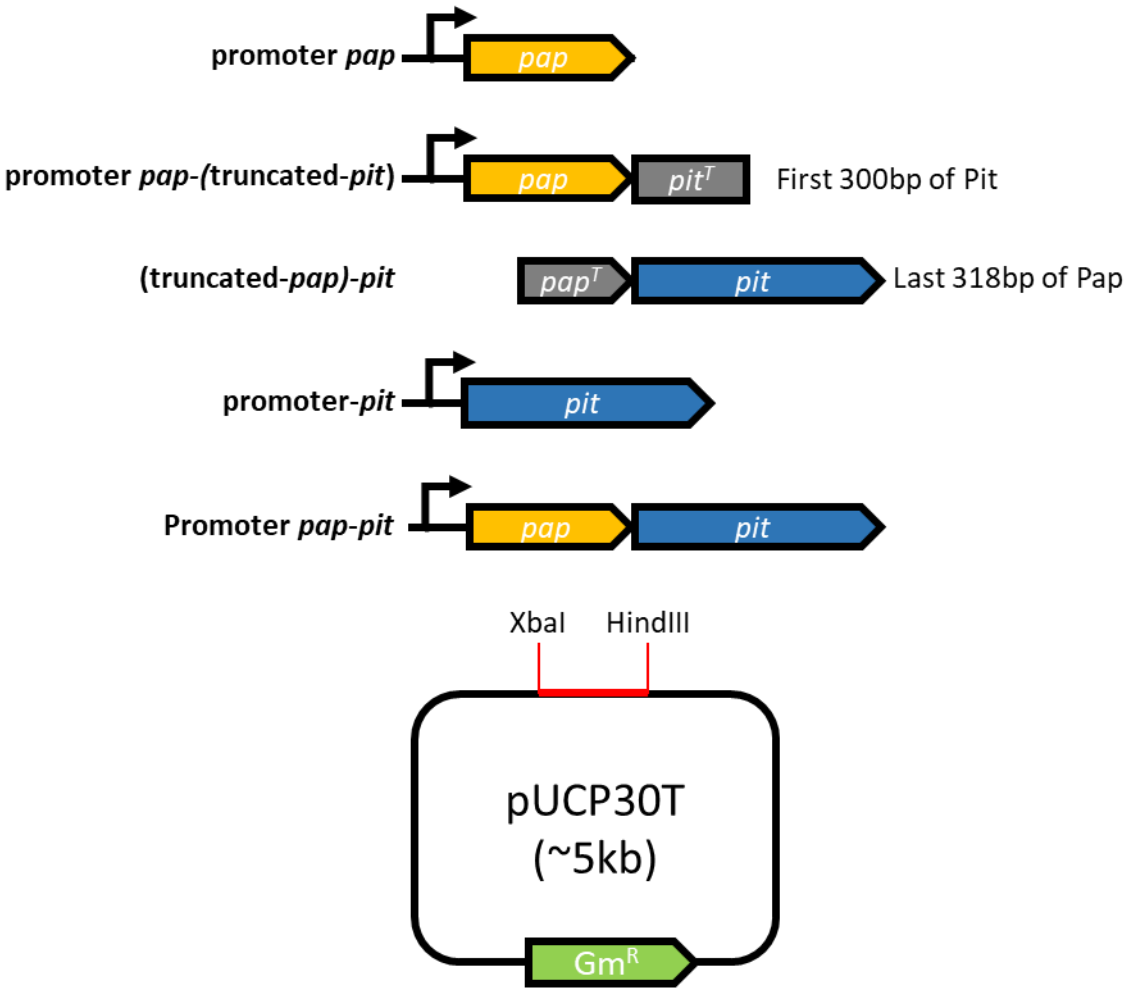


Figure 2-6. Diagram of the *pap-pit* loci cloned into pUCP30T used for complementation by genome recombination. *S. meliloti pap-pit* deletion mutants were complemented by single-recombination of a pUCP30T vector containing one of five *pap-pit* regions PCR amplified and cloned by ligation via XbaI/HindIII. The promoter region included in the constructs includes 250bp upstream of the *pap* start codon.

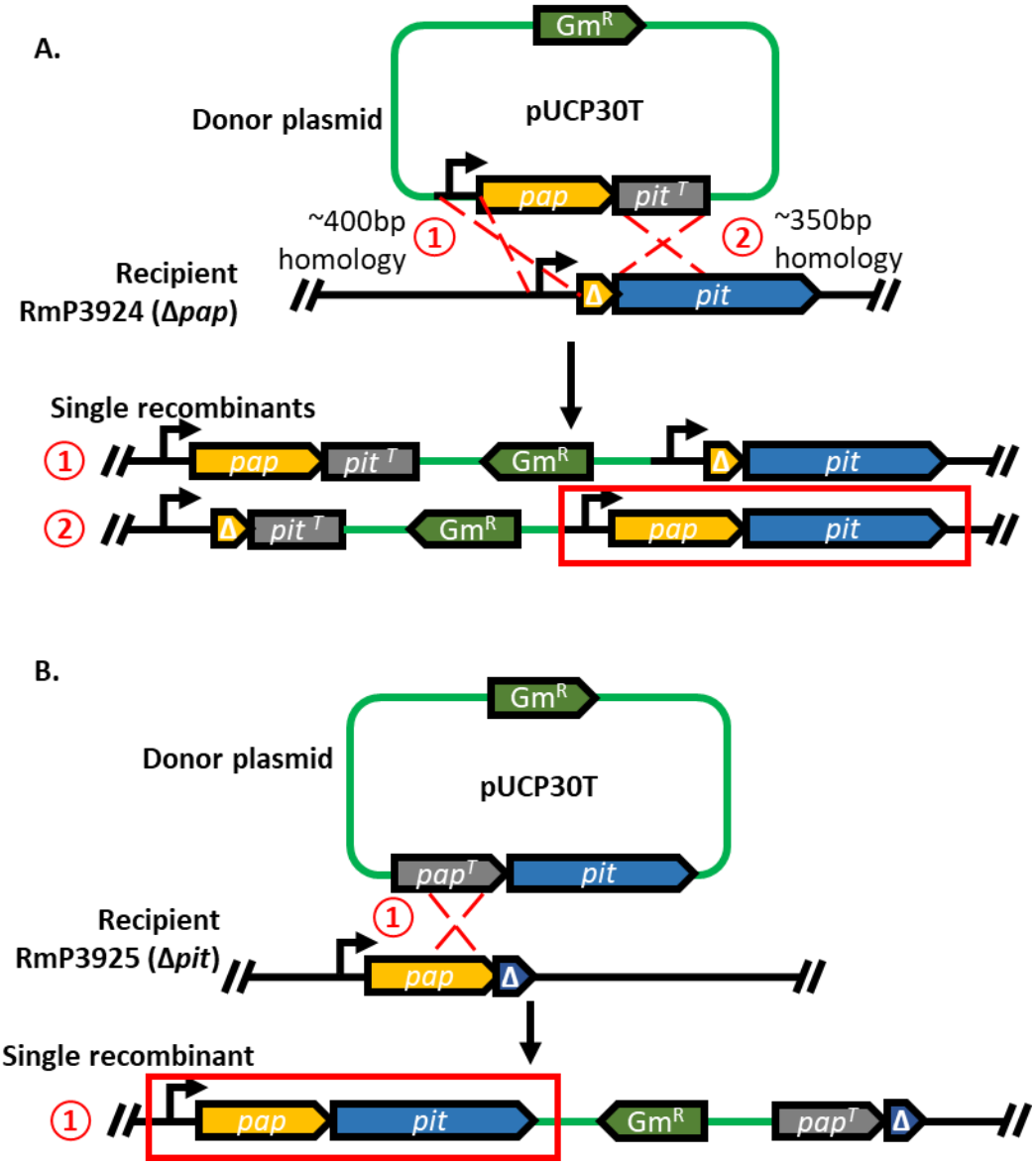


Figure 2-7. Diagram of recombinants produced by integration of plasmids into *pap* and *pit* deletion mutants. pUCP30T plasmids were generated with fragments of the *pap-pit* operon such that recombination would result in integration of the fragment within the *pap* or *pit* deletion region. **A.** Δ*pap* recombination with *pap-pit*^{truncated} was expected to result in two different recombinants, where recombination would occur in the promoter region (1) or between the truncated *pit* region (2). Only recombinant (2) would result in regeneration of the *pap-pit* operon structure. **B.** Δ*pit* recombination with *pap*^{truncated}*-pit* was expected to result in a single recombination between the *pap* regions, resulting in integration of the wildtype *pit* gene downstream of the *pap* gene on the chromosome.

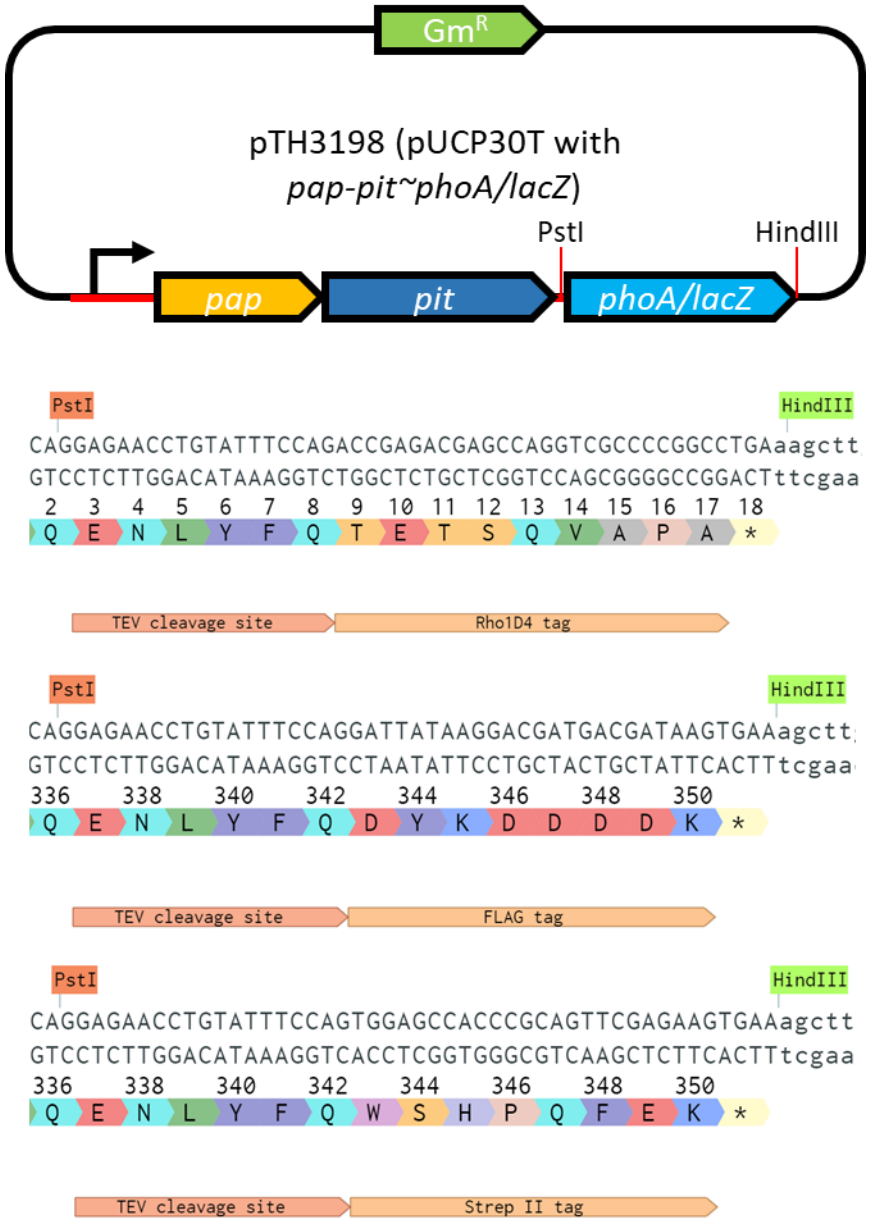


Figure 2-8. Diagram of the recombinant peptide tags generated to the C-terminal end of the SmPit protein. Recombinant SmPit~peptide tag constructs were generated by cloning small oligos (~60-70 bp) into pTH3198 (pUCP30T carrying *pap-pit~phoA/lacZ*) via PstI and HindIII restriction sites, which directly flank the *phoA/lacZ* reporter fusion. The resulting clones have an 8 a.a. linker region between the Pit coding sequence and the tag peptide. These vectors were subsequently conjugated into RmP3926 (Δ*pap-pit*) to examine whether the recombinant Pit~tag fusions could be detected by Western Blot.

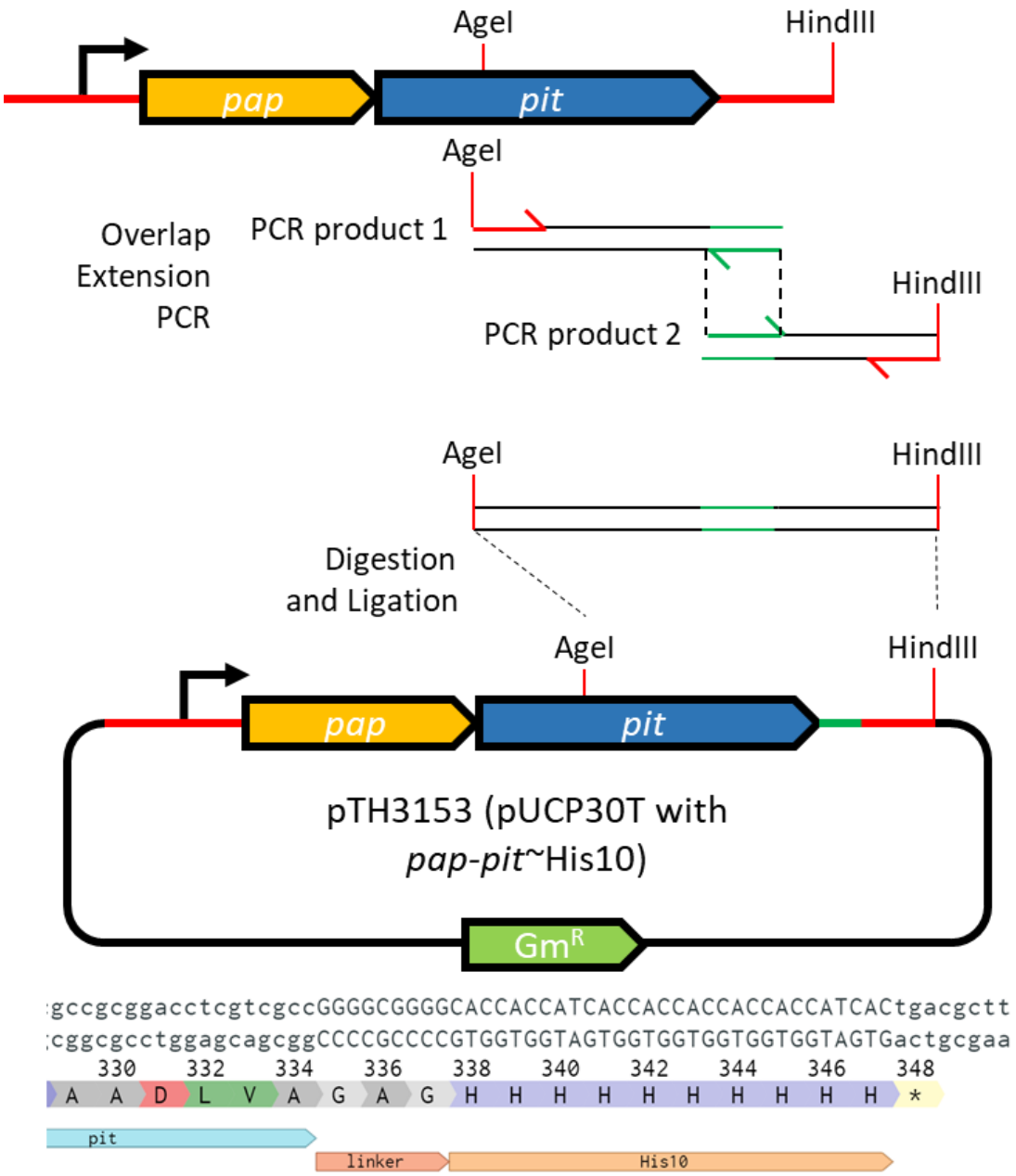


Figure 2-9. Diagram of generating the SmPit~His10 tag by overlap extension PCR. A 10X-polyhistidine tag (His10) fusion to SmPit was also generated. Two PCR products with a 40bp homology were PCR amplified, with the homology region containing the sequence of the His10 tag. The PCR products were then joined by overlap extension PCR and subsequently cloned into pTH2824 (pUCP30T - *wt pap-pit*) via AgeI and HindIII restriction sites.

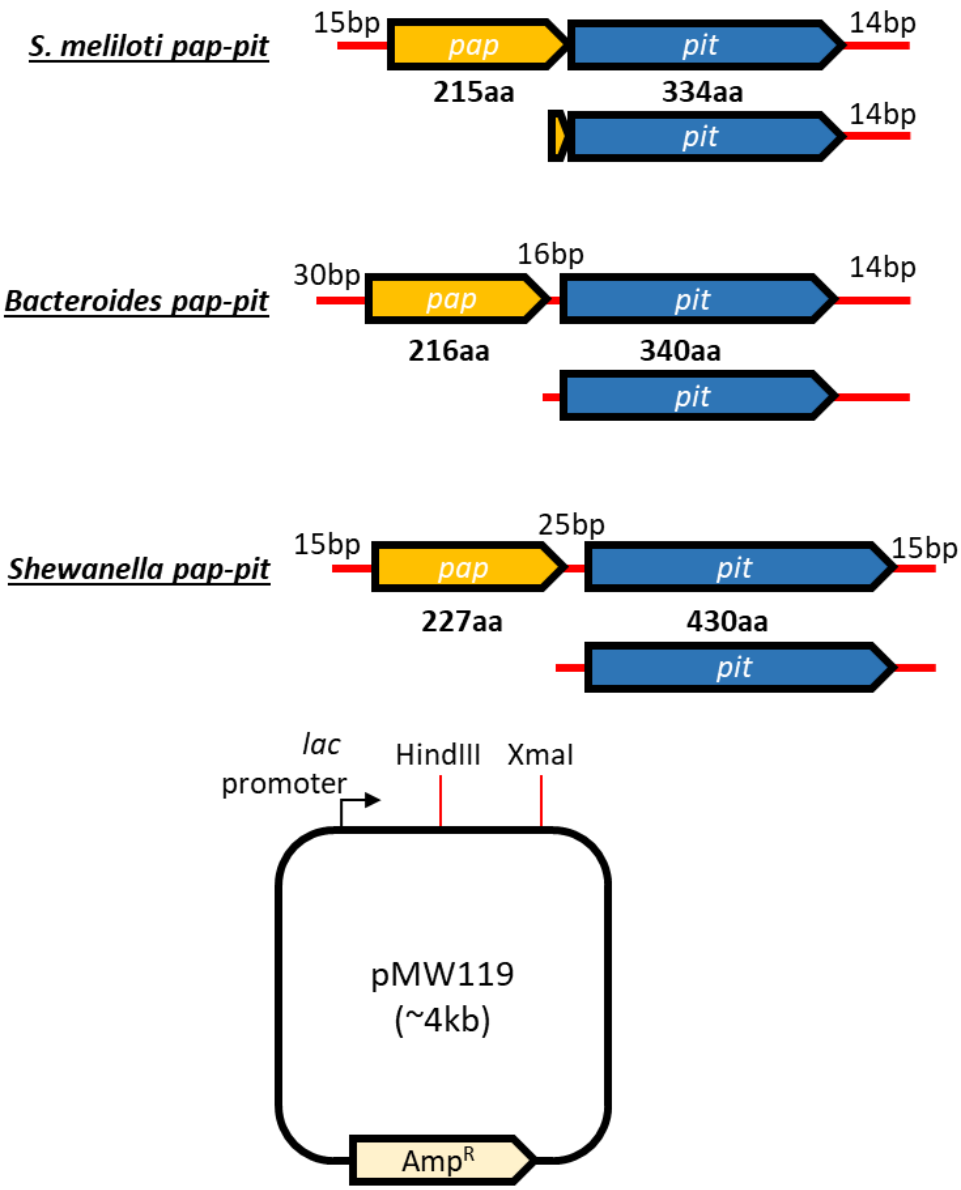
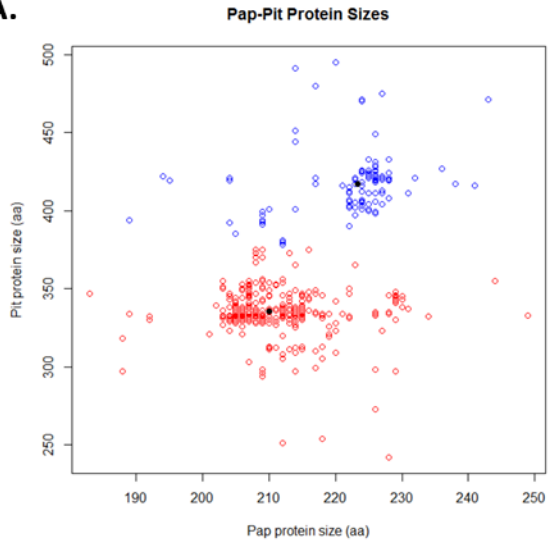


Figure 2-10. Diagram of the heterologous *pap-pit* operons cloned into pMW119. Pap-Pit orthologues from *S. meliloti* RmP110, *Bacteroides thetaiotaomicron* VPI-5482, and *Shewanella oneidensis* MR-1 were PCR amplified and cloned into the low copy-number vector, pMW119 via HindIII and XmaI under control of the *lac* promoter. The regions upstream of *pap* were also amplified to include the native ribosome binding sites. Additionally, the *pit* genes alone were also cloned, including some region upstream for a putative ribosome binding site.

Domain	Class	Pap-Pit Count	Pap-Pit (%)	Pit-only Count	Pit-only (%)	Difference (%)
Archaea	Bathyarchaeota	1	0.11%	0	0.00%	0.11%
Archaea	Crenarchaeota	9	1.03%	11	0.95%	0.07%
Archaea	Euryarchaeota	82	9.35%	178	15.42%	**-6.07%**
Archaea	Lokiarchaeota	0	0.00%	2	0.17%	-0.17%
Archaea	Thaumarchaeota	8	0.91%	1	0.09%	0.83%
Archaea	UnclassifiedArchaea	0	0.00%	4	0.35%	-0.35%
Bacteria	Acidobacteria	5	0.57%	6	0.52%	0.05%
Bacteria	Actinobacteria	99	11.29%	215	18.63%	**-7.34%**
Bacteria	Alphaproteobacteria	62	7.07%	114	9.88%	**-2.81%**
Bacteria	Aquificae	1	0.11%	0	0.00%	0.11%
Bacteria	Armatimonadetes	1	0.11%	0	0.00%	0.11%
Bacteria	Bacteroidetes	42	4.79%	100	8.67%	**-3.88%**
Bacteria	Betaproteobacteria	93	10.60%	59	5.11%	**5.49%**
Bacteria	Caldiserica	1	0.11%	0	0.00%	0.11%
Bacteria	Chlamydiae	16	1.82%	1	0.09%	1.74%
Bacteria	Chlorobi	1	0.11%	2	0.17%	-0.06%
Bacteria	Chloroflexi	5	0.57%	0	0.00%	0.57%
Bacteria	Chrysiogenetes	1	0.11%	0	0.00%	0.11%
Bacteria	Cyanobacteria	3	0.34%	32	2.77%	**-2.43%**
Bacteria	Deferribacteres	1	0.11%	0	0.00%	0.11%
Bacteria	Deinococcus-Thermus	11	1.25%	0	0.00%	1.25%
Bacteria	Deltaproteobacteria	49	5.59%	15	1.30%	**4.29%**
Bacteria	Elusimicrobia	1	0.11%	0	0.00%	0.11%
Bacteria	Epsilonproteobacteria	1	0.11%	46	3.99%	-3.87%
Bacteria	Fibrobacteres	0	0.00%	1	0.09%	-0.09%
Bacteria	Firmicutes-Bacilli	107	12.20%	38	3.29%	**8.91%**
Bacteria	Firmicutes-Clostridia	68	7.75%	5	0.43%	**7.32%**
Bacteria	Firmicutes-Others	12	1.37%	3	0.26%	1.11%
Bacteria	Gammaproteobacteria-Enterobacteria	7	0.80%	142	12.31%	**-11.51%**
Bacteria	Gammaproteobacteria-Others	153	17.45%	137	11.87%	**5.57%**
Bacteria	Gemmatimonadetes	2	0.23%	4	0.35%	-0.12%
Bacteria	Nitrospirae	3	0.34%	6	0.52%	-0.18%
Bacteria	Planctomycetes	1	0.11%	8	0.69%	-0.58%
Bacteria	Spirochaetes	3	0.34%	0	0.00%	0.34%
Bacteria	Thermodesulfobacteria	2	0.23%	0	0.00%	0.23%
Bacteria	Thermotogae	21	2.39%	20	1.73%	0.66%
Bacteria	UnclassifiedBacteria	1	0.11%	0	0.00%	0.11%
Bacteria	UnclassifiedTerrabacteriagroup	1	0.11%	1	0.09%	0.03%
Bacteria	Verrucomicrobia	3	0.34%	3	0.26%	0.08%

Table 3-1. Table of Pit systems by phylogenetic class. Pap-Pit and Pit-only orthologue groups were sorted by phylogenetic domain and class. The percentage of orthologues (%) was determined by the quotient of the count from the total count within the group. The difference (%) is the percentage difference between the Pap-Pit and Pit-only groups.

A.



Cluster	Mean (a.a.)		Count
	Pap	Pit	
A	212	332	639
B	222	420	238

B.

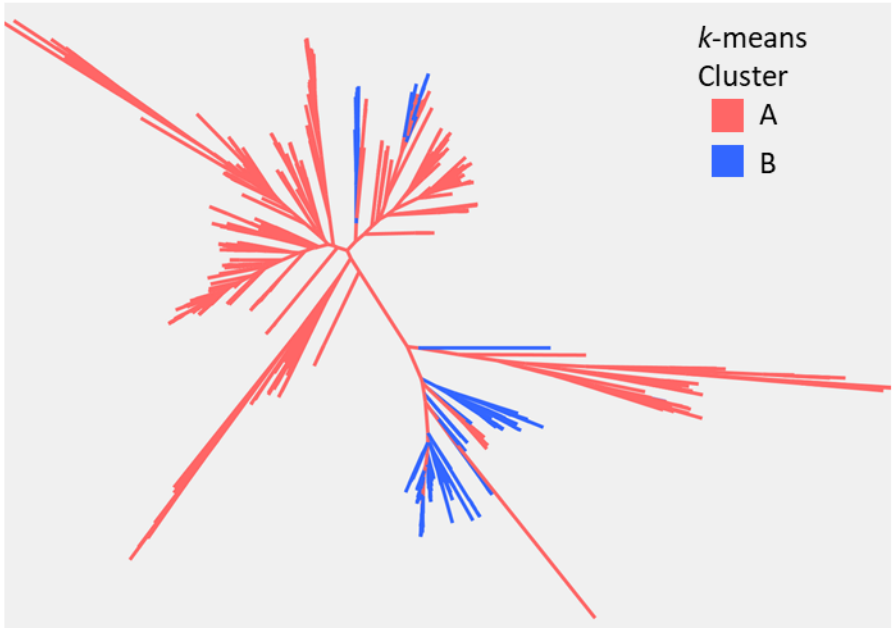


Figure 3-2. *k*-means clustering of Pap and Pit proteins and phylogeny of Pap-associated Pit proteins. *k*-means clustering was used to predict that there are 2 significant clusters of Pap-Pit protein pairs. **A.** Pap-Pit orthologues were plotted with the a.a. size of the proteins as the X-Y axis, with distinct mean protein sizes. **B.** Maximum likelihood phylogeny of concatenated Pap-Pit pairs shows that these clusters are relatively accurate, however the distinctions are likely not solely due to size.

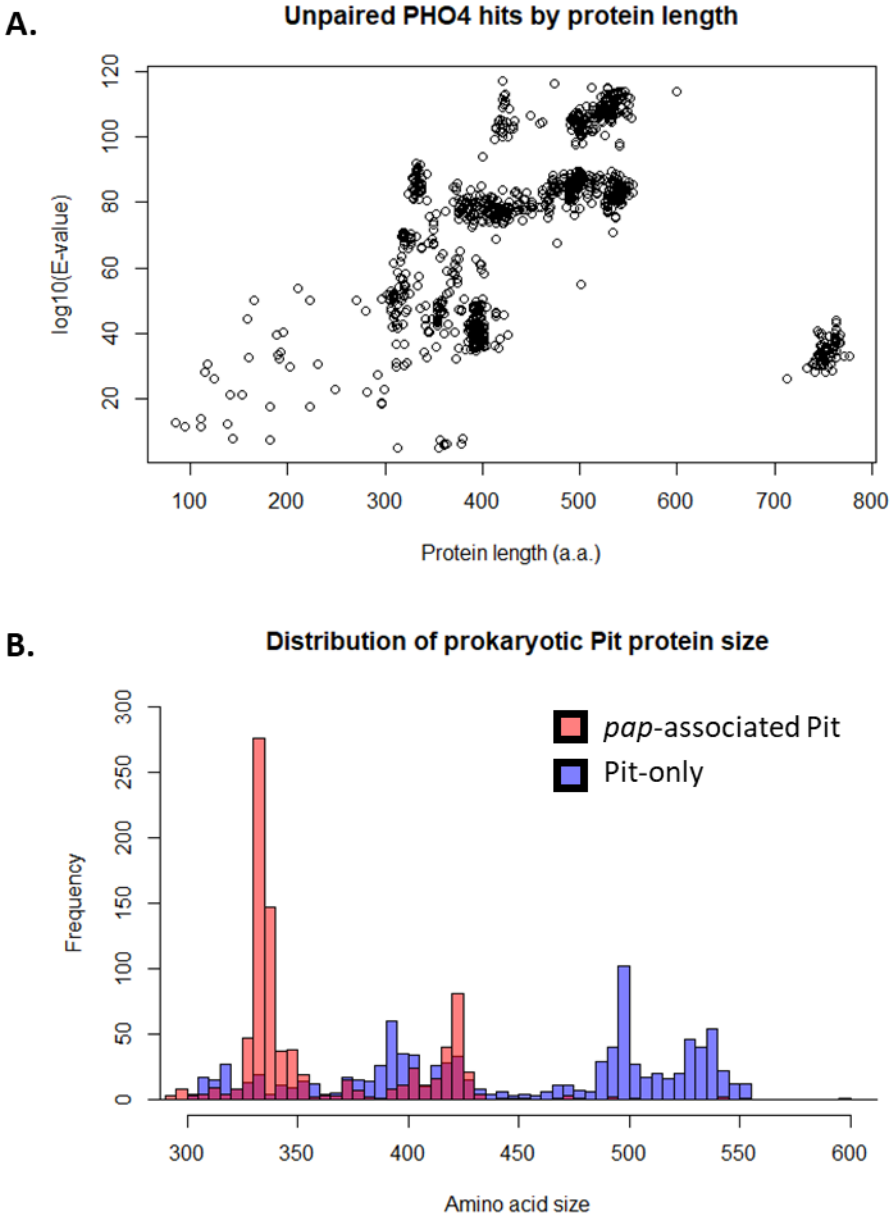


Figure 3-4. Comparison of Pit-only and Pap-associated Pit proteins by protein size. Unpaired Pit proteins (Pit-only) were identified as PHO4 best-hits not found adjacent to a Pap orthologue. **A**. A plot of the Pit-only protein size (a.a.) versus the HMM log10 E-value score, shows that there are several clusters of Pit-only proteins. **B.** Histogram comparing *pap*-associated Pit (red) and Pit-only (blue) proteins by protein size (a.a.). Show *pap*-associated Pit proteins encode proteins of smaller size.

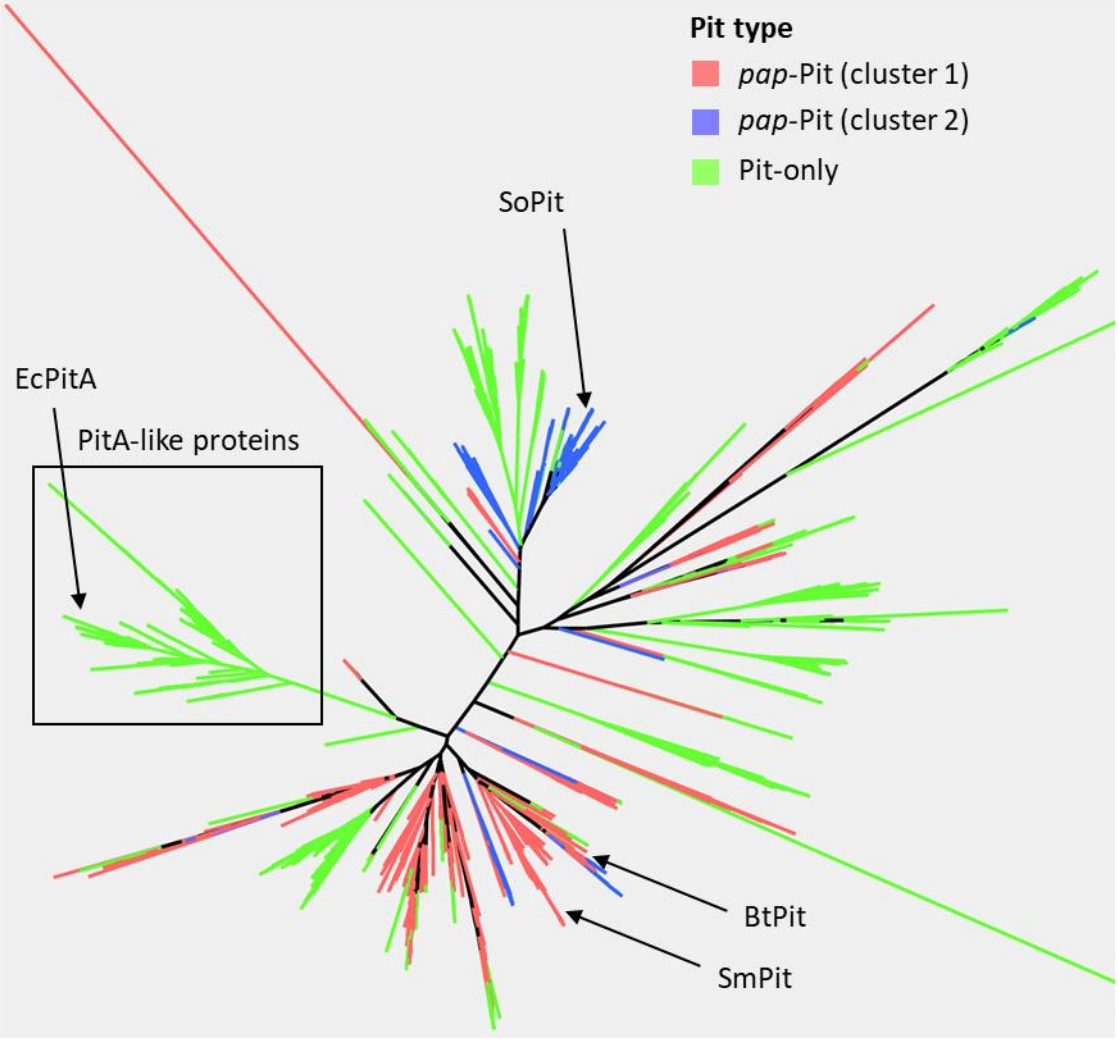


Figure 3-5. Phylogeny of Pit proteins (Pit-only and Pap-associated Pit). Maximum likelihood phylogeny of *pap-associated* Pit (red and blue) and Pit-only proteins (green) show that Pit-only proteins do not form a distinct clade from *pap*-Pit proteins. Rather, majority of the proteins are more closely related to *pap*-Pit, with the exception of two clades: one of which includes *E. coli* PitA and PitA-like proteins, and a second that is found within the *pap*-Pit minor cluster.

Phylogenetic Domain	Count	(%)
Acidobacteria	6	1.8%
Actinobacteria	1	0.3%
Alphaproteobacteria	37	11.3%
Bacteroidetes	16	4.9%
Betaproteobacteria	44	13.4%
Cyanobacteria	2	0.6%
Deltaproteobacteria	2	0.6%
Epsilonproteobacteria	2	0.6%
Euryarchaeota	1	0.3%
Gammaproteobacteria – Enterobacteria	142	**43.4%**
Gammaproteobacteria – Others	72	**22.0%**
Planctomycetes	2	0.6%

Figure 3-6. Phylogenetic distribution of PitA-like proteins. Maximum likelihood phylogeny of PitA-like proteins coloured by the phylogenetic domain from which each protein sequence originates. 91% of the sequences originate from proteobacteria, however a large majority derive from Gammaproteobacteria (>65%). There also appears to be a distinct clade within this group that is composed of 100% Gammaproteobacteria that includes EcPitA.

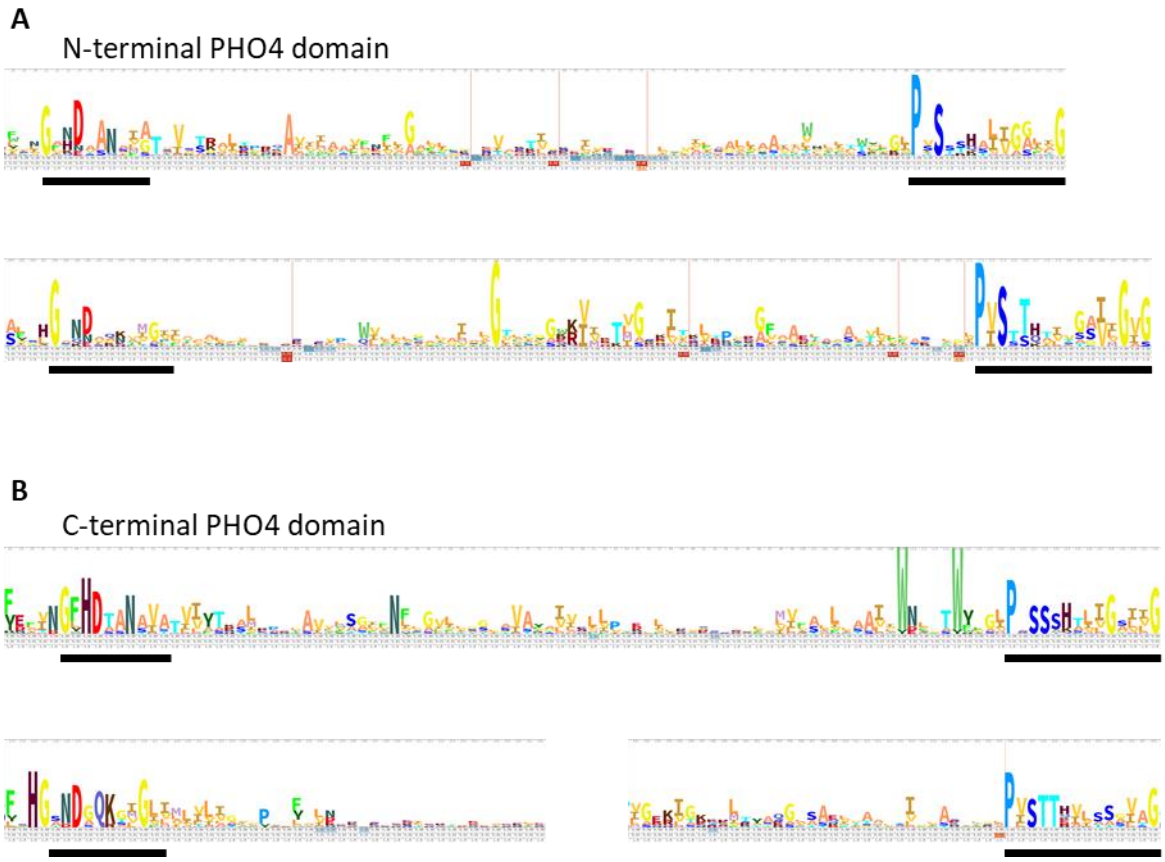


Figure 3-7. HMM logo of the PHO4 domains in Pit proteins. HMM logo of the PHO4 domains found in *pap*-associated Pit (**A**) and PitA-like (**B**) protein. Each Pit protein has two PHO4 signature sequence motifs of GXNDXAN - PVS(S/T)$_3$HXXG, with a poorly conserved region in between of 70-75 a.a.

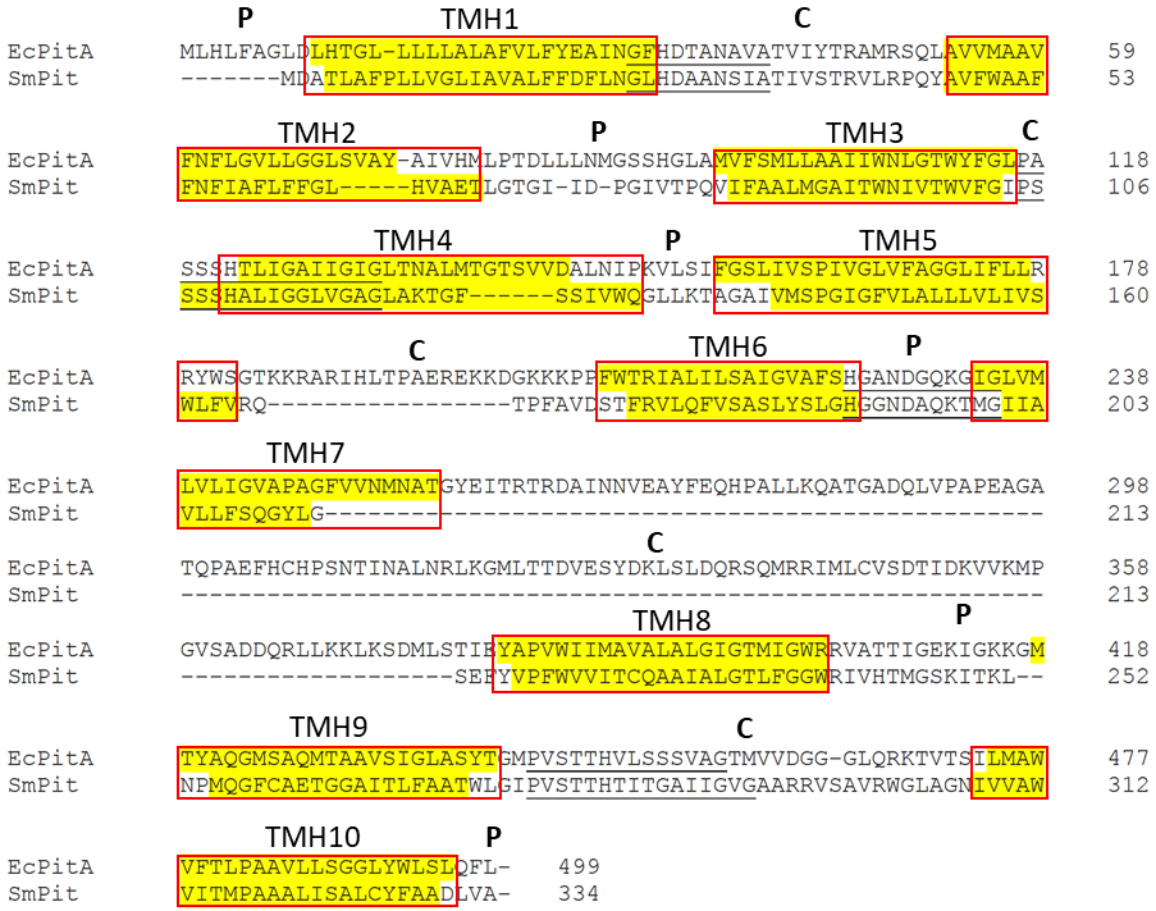


Figure 3-8. Alignment of the SmPit and EcPitA protein sequences and membrane topology. ClustalW alignment and TMHMM transmembrane helices (TMH) prediction of SmPit and EcPitA indicate that both have 10 TMH that appear to be relatively conserved. The N- and C-terminal ends are predicted to be within the periplasm (P), however EcPitA has a 122 a.a. loop within the cytoplasm (C). The intracellular loop is present within the C-terminal PHO4 signature sequences of EcPitA.

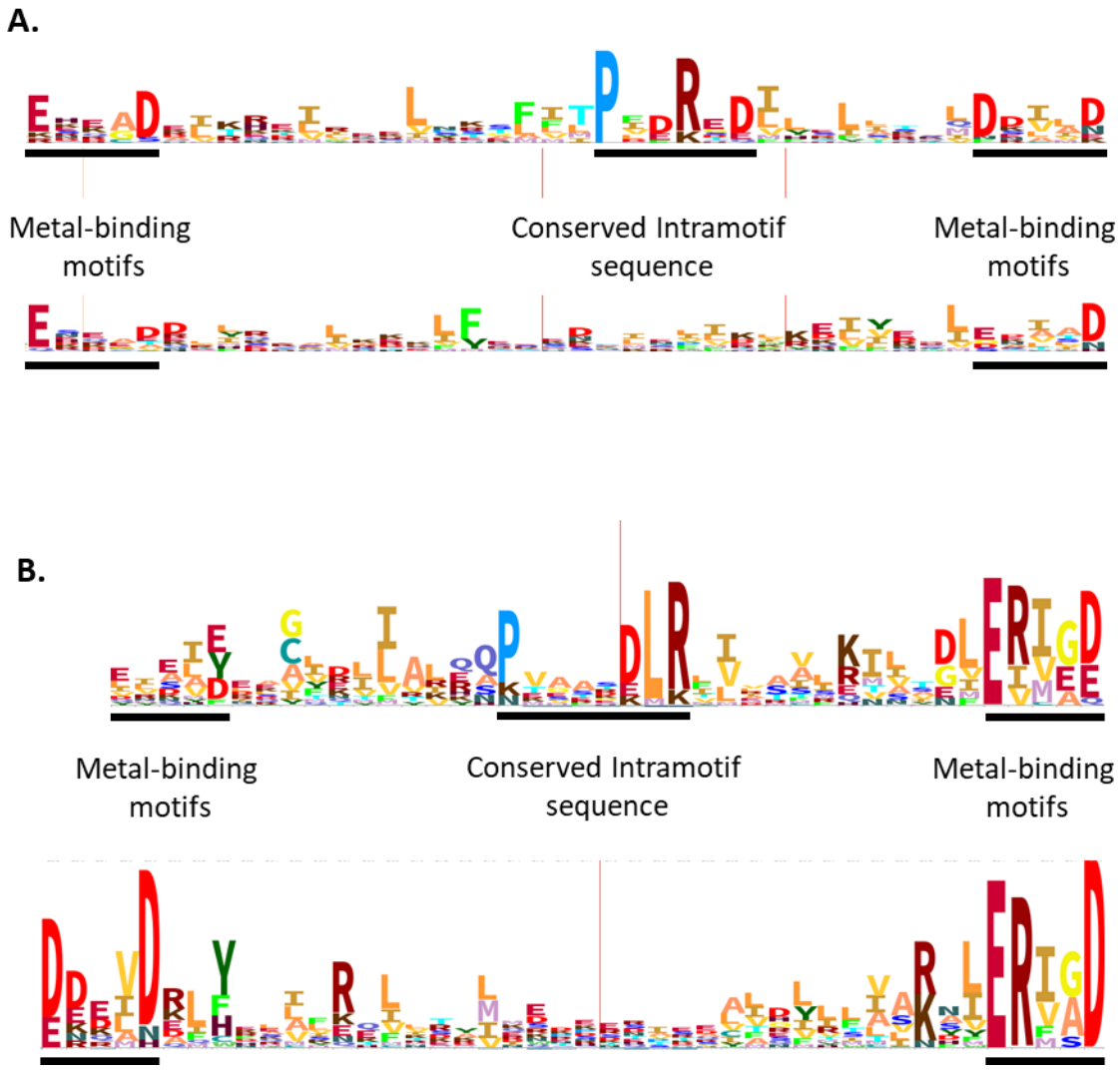


Figure 3-9. HMM logo of Pap and PhoU proteins. **A.** HMM logo of Pap proteins and **B.** PhoU proteins. Both proteins have two PhoU domains, each with two conserved PhoU motif signature sequences. However, the PhoU motif signature sequences are differentially conserved between Pap and PhoU proteins, particularly the second PhoU-motifs in PhoU proteins appear to be more highly conserved (ERXXD), whereas the Pap motif is less conserved. Also, the highly conserved region between PhoU-motifs in the N-terminal PhoU domain is PXXRXD for Pap proteins, but PXXXXDLR for PhoU proteins.

Figure 3-10. Alignment of Pap orthologues and PhoU highlighting conserved regions. Alignment of SmPhoU, SoPap, SmPap and BtPap with highlighted PhoU-like motifs (yellow) and conserved intramotif region (blue). Single missense mutations were generated for residues of *S. meliloti* Pap (bolded letters) and labeled below.

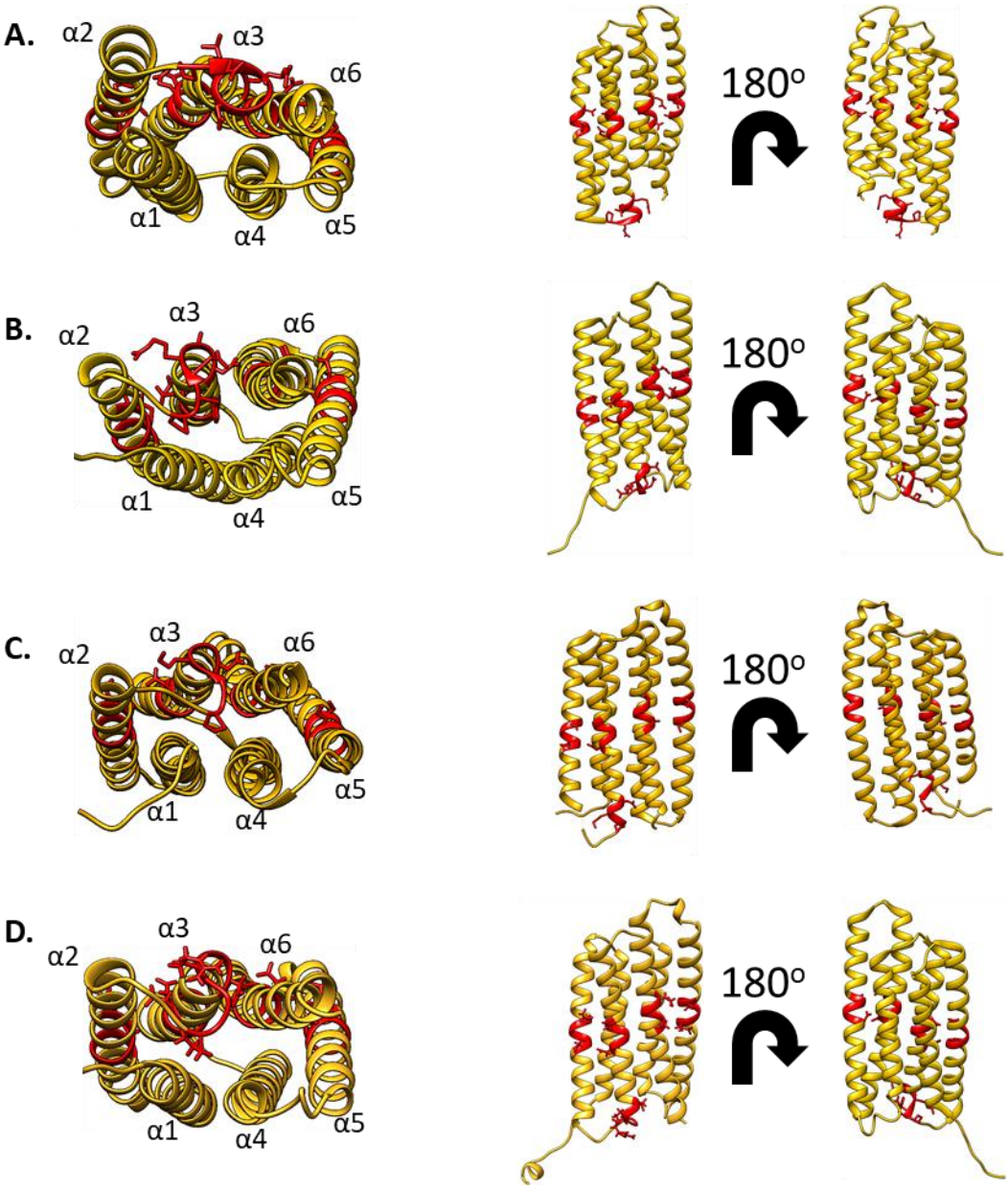


Figure 3-11. Structural comparison of Pap orthologues and PhoU. The PhoU-like motifs (red) of SmPhoU (A), BtPap (B), SoPap (C), and SmPap (D) are all present on α2-α3 and α5-α6 and along a perpendicular line. The highly conserved region between the first two PhoU motifs is also found between α2-α3 in all structures.

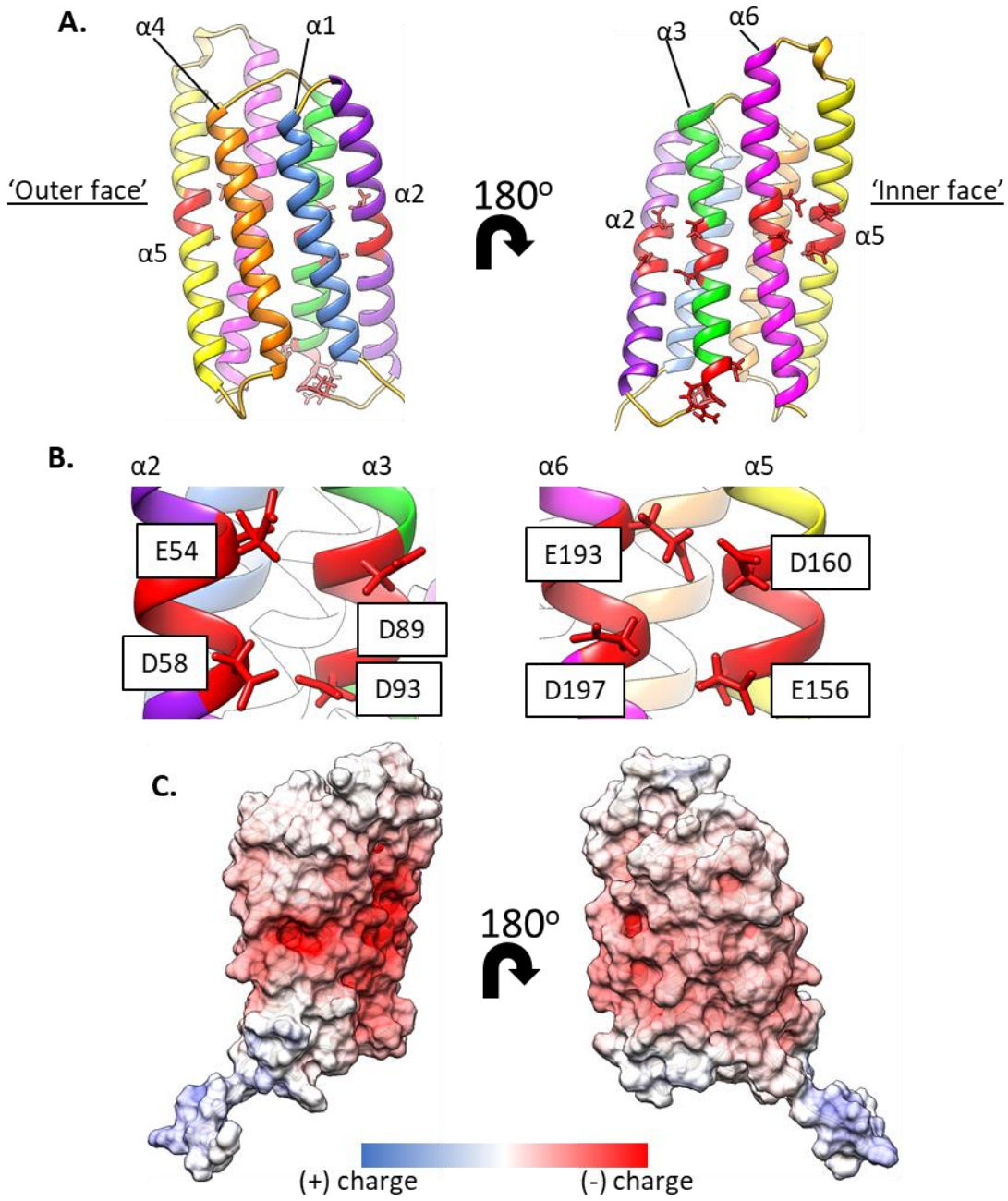


Figure 3-12. I-TASSER Structural prediction of S. meliloti Pap. I-TASSER predicts SmPap to form a bundle of six parallel α-helices separated by short loop regions. **A.** PhoU-motifs (red) are found on α2, α3, α5 and α6, whereas the highly conserved intramotif region (PXXRXD) is present between α2-α3, similar to that found in PhoU. **B.** The PhoU-motifs are found on residues perpendicular to the α-helices, which may form a binding pocket for a ligand or other protein. C. The electrostatic potential of SmPap shows a highly positive charge on the 'inner face' and a neutral 'outer face'.

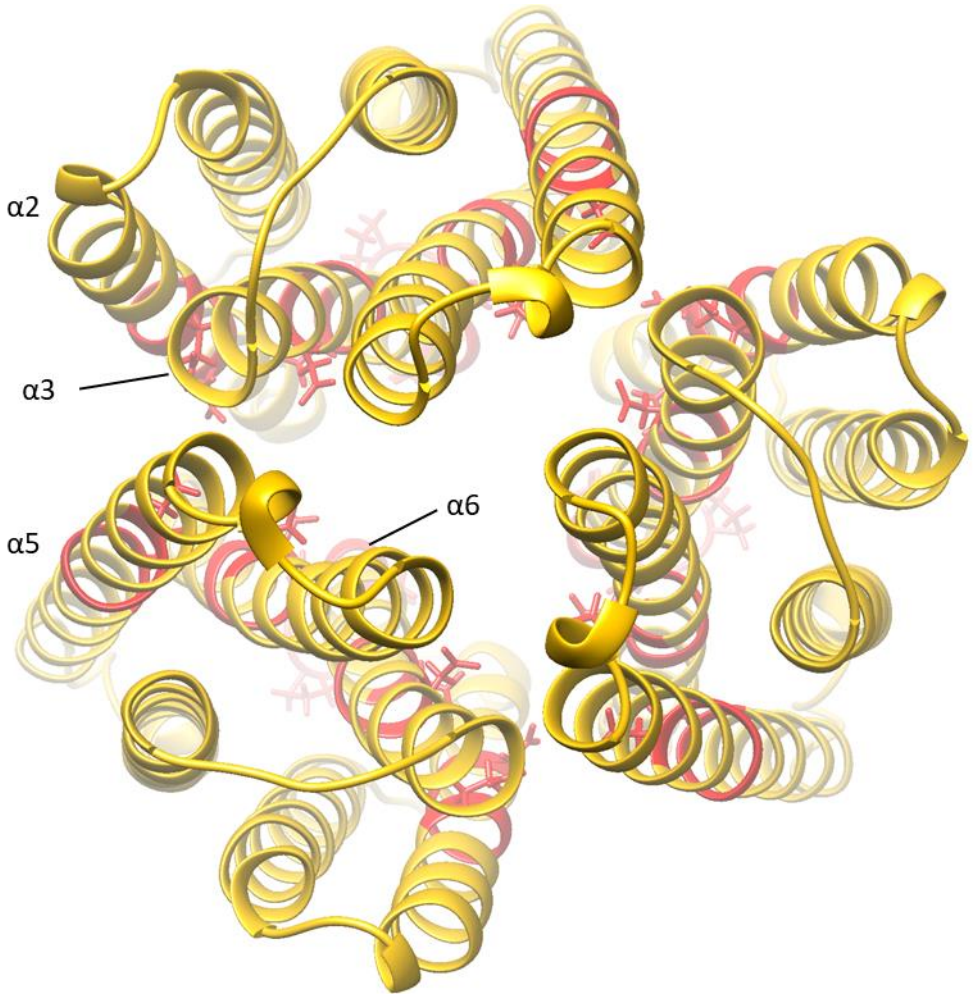


Figure 3-13. Putative trimeric structure of SmPap. SmPap proteins in the may possibly form a homotrimer structure, as predicted by the crystal structure of BtPap (PDB: 3L39). α2-α3 and α5-α6 are parallel and the PhoU-like motif residues are directly adjacent between two SmPap monomers. This may indicate that a Pap-Pap interaction is dependent on these the PhoU-like motifs.

Table 3-14A – Other Pap hits

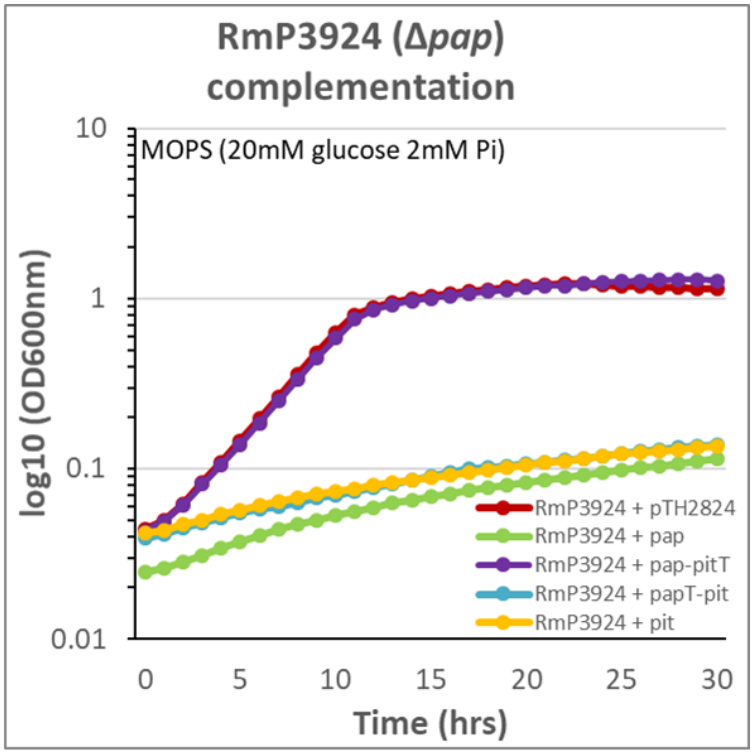
Pfam Motif	Mean E-value	Count (# of hits)
Na_Pi_cotrans	E-04	63
MCPsignal	E-02	47
TrkA_C	E-03	37
HATPase_c	E-03	11
MFS_1	E-17	4

Table 3-14B – Other PhoU hits

Pfam Motif	Mean E-value	Count (# of hits)
Na_Pi_cotrans	E-03	532
PhoU_div	E-02	377
TrkA_C	E-04	123
MazE_antitoxin	E-3	14

Table 3-14. Table of other best-hits to proteins with PhoU or PhoU_div domains. Other best hits were identified as PhoU or PhoU_div hits, where the most significant E-value was another Pfam domain. Both PhoU and PhoU_div other hits were highly associated with Na_Pi_cotrans domains, which are a poorly studied class of membrane proteins in bacteria. Additional domains were identified which do not have any clear connection to Pi metabolism or homeostasis.

A.



B.

Strain	Velocity (nmol Pi mg^{-1} min^{-1})
RmP3924 + pTH2824	3.13[a]
RmP3924 + *pap*	0.78[b]
RmP3924 + *pap-pit*^{T}	3.48[a]
RmP3924 + *pap*^{T}*-pit*	0.07[b]
RmP3924 + *pit*	0.14[b]

Figure 3-15. Complementation of *S. meliloti* Δ*pap* mutant by integration of the wildtype *pap* gene into the deletion region. A. The growth defect of RmP3924 (Δ*pap*) in MOPS-P2 was only complemented by recombination of pUCP30T carrying *pap-pit*^{T} and wildtype *pap-pit* into the chromosome. **B.** Similarly, Pi uptake velocity of RmP3924 (*pap-pit*^{T}) and was observed to be similar to RmP3924 (*pap-pit*), which indicates that Pi uptake had been restored to the mutant. Statistical groupings are indicated by superscript letters determined by one-way ANOVA followed by a Tukey's post-hoc test ($p < 0.05$).

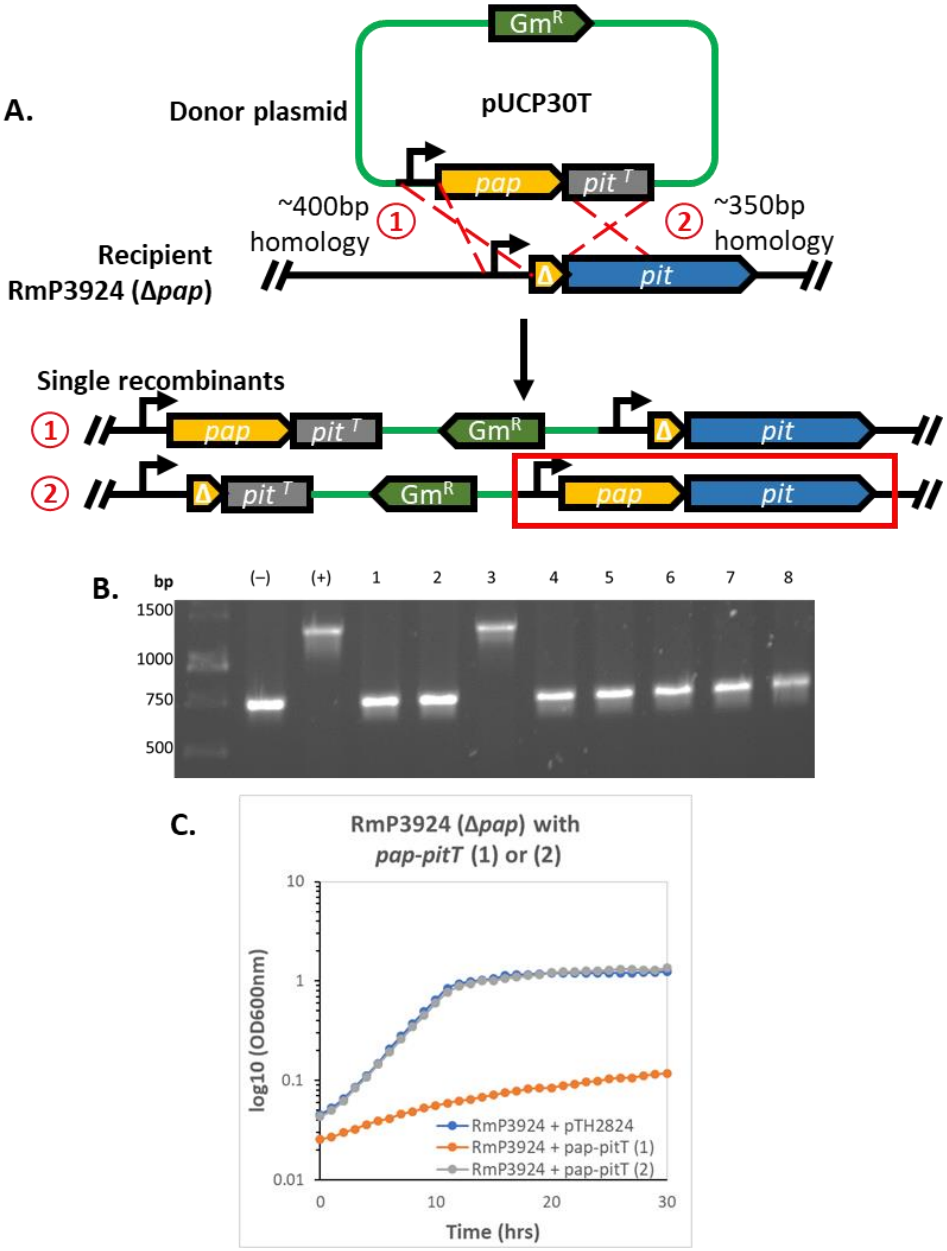
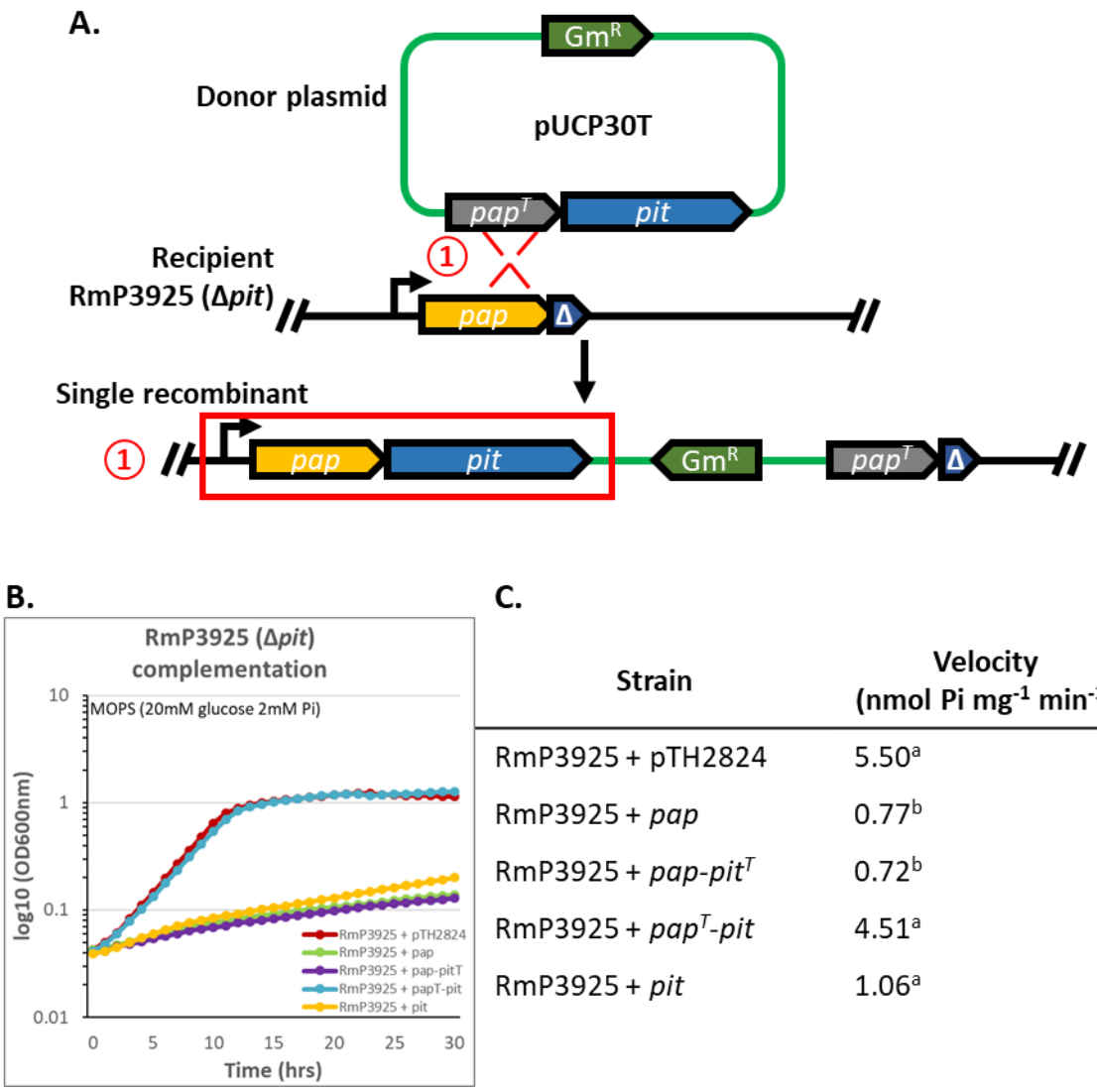


Figure 3-16. Two possible recombinants by *pap-pit*^{truncated} integration. **A.** Single cross-over recombination of pUCP30T (*pap-pit*^{T}) into RmP3924 resulted in two recombinants, where (1) *pap* and *pit* are under separate promoters or (2) *pap-pit* are encoded within an operon. **B.** Recombinants (1) and (2) could be discerned by colony PCR by amplifying the *pap-pit* operon. (-) RmP3924 negative control, (+) RmP3924 (pTH2824) positive control. **C.** Recombinant (2) was also demonstrated to complement RmP3924, whereas recombinant (1) was not, as compared with RmP3924 (pTH2824).



Strain	Velocity (nmol Pi mg^{-1} min^{-1})
RmP3925 + pTH2824	5.50[a]
RmP3925 + *pap*	0.77[b]
RmP3925 + *pap-pit*^{T}	0.72[b]
RmP3925 + *pap*^{T}*-pit*	4.51[a]
RmP3925 + *pit*	1.06[a]

Figure 3-17. Complementation of *S. meliloti* Δ*pit* by integration of the wildtype *pit* gene into the deletion region. A. A single cross-over recombination of RmP3925 (Δ*pit*) with pUCP30T (*pap*^{T}*-pit*) resulted in one recombinant, where the *pit* gene from the plasmid was integrated downstream of *pap* on the chromosome. **B.** The growth defect of RmP3925 in MOPS-P2 was only complemented by recombination of pUCP30T(*pap*^{T}*-pit*) or pTH2824 (*pap-pit*), but not pUCP30T (*pit*). **C.** This observation was reflected in the Pi uptake velocity, as RmP3925 (*pap*^{T}*-pit*) had similar velocity as RmP3925 (pTH2824). Statistical groupings are indicated by superscript letters determined by one-way ANOVA followed by a Tukey's post-hoc test ($p < 0.05$).

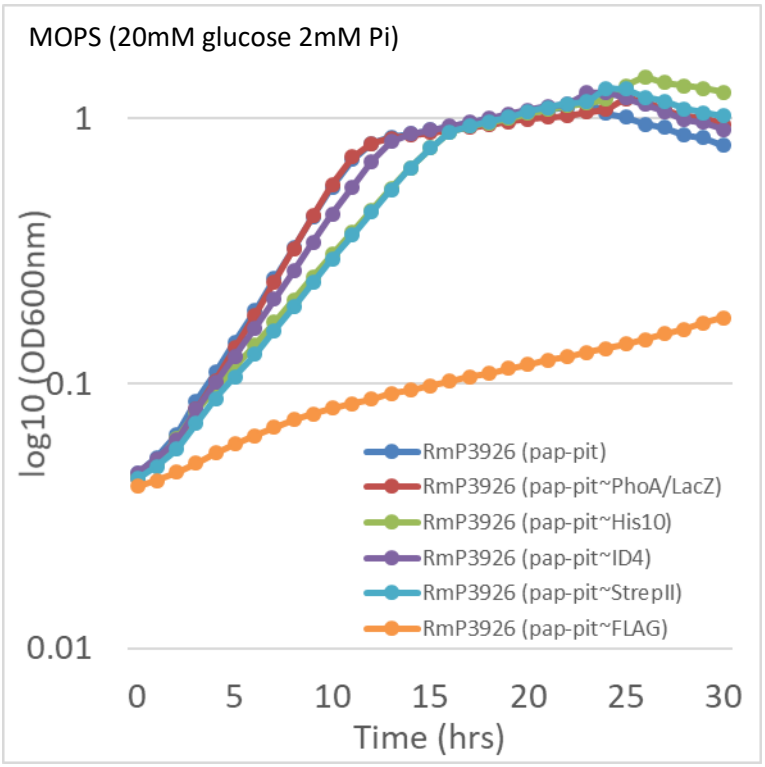


Figure 3-18. Complementation of *S. meliloti* Δ*pap-pit* by *pap-pit*~Tag recombinant constructs. Growth was observed for all *pap-pit*~Tag constructs similar to the wildtype *pap-pit* in MOPS-P2 except for *pap-pit*~FLAG. This indicates that the recombinant tags (except the FLAG tag) do not affect the transport function of the Pap-Pit system.

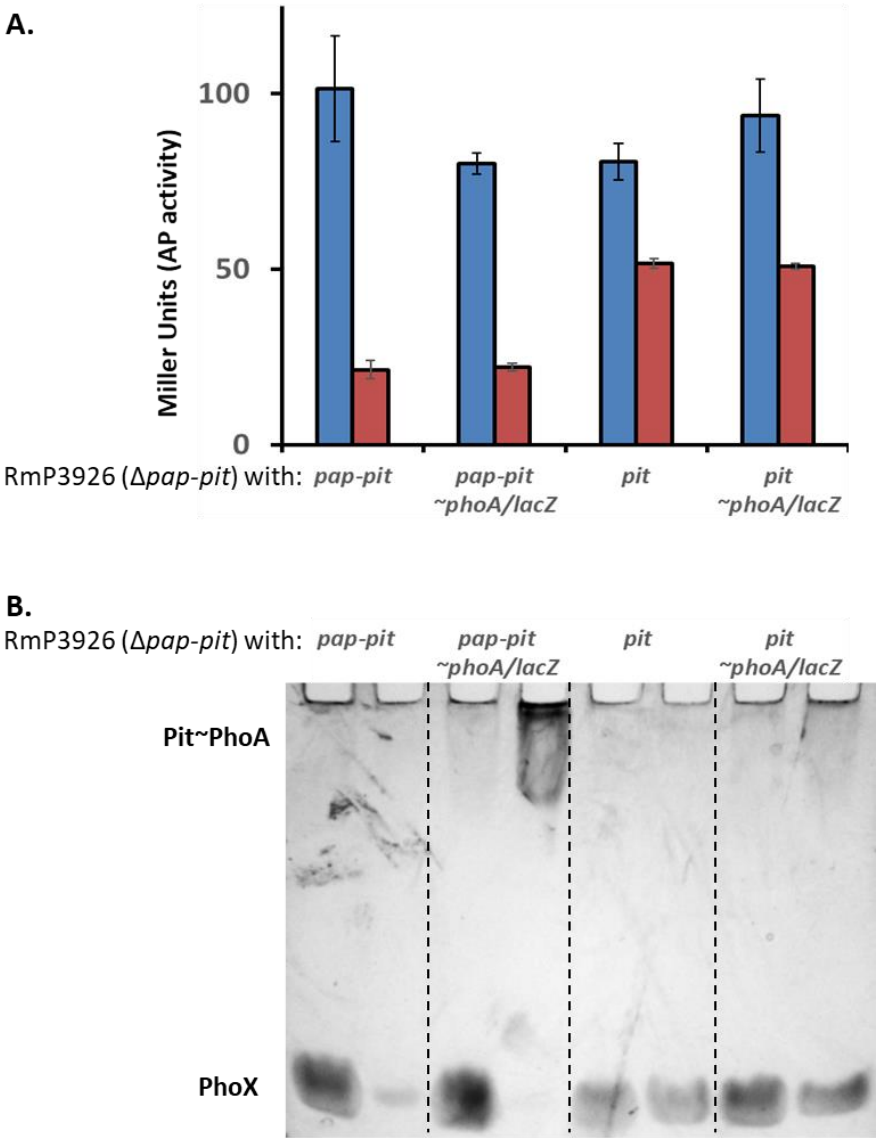


Figure 3-19. Alkaline phosphatase activity and in-gel zymogram of Pit~PhoA/LacZ recombinant tag. A. AP activity of *S. meliloti* incubated in MOPS-P0 (blue) or MOPS-P2 (red) for 16 hours, shows no significant difference in activity between RmP3926 (Δ*pap-pit*) with wildtype *pap-pit* or *pap-pit~phoA/lacZ*. There is also no significant difference between RmP3926 with *pit* or *pit~phoA/lacZ*. **B.** In-gel zymogram of AP activity from the same recombinants, separated by cytoplasmic and membrane fractions, respectively. AP activity is detected only in the membrane fraction of RmP3926 with *pap-pit~phoA/lacZ*, but not *pap-pit*.

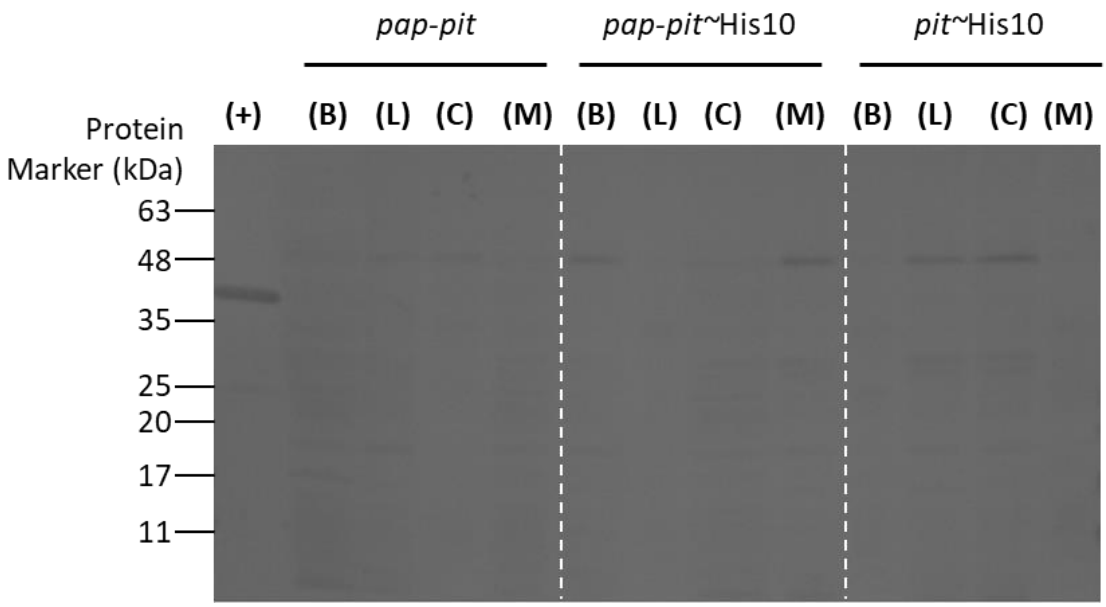


Figure 3-20. Western blot of *S. meliloti* Pit~His10 tag. Anti-His western blot of RmP3926 with either *pap-pit*, *pap-pit*~His10, *pit*~His10 demonstrated no distinct bands present in the samples containing the His-tagged SmPit protein. For each sample, a boiled lysate (B), sonicated lysate (L), cytoplasmic fraction (C), and membrane fraction (M) were run on the SDS-PAGE, respectively. (+) PckR-His6 was used as a positive control to confirm the anti-His antibodies were present. Although there appears to be a band at approximately 40kDa.

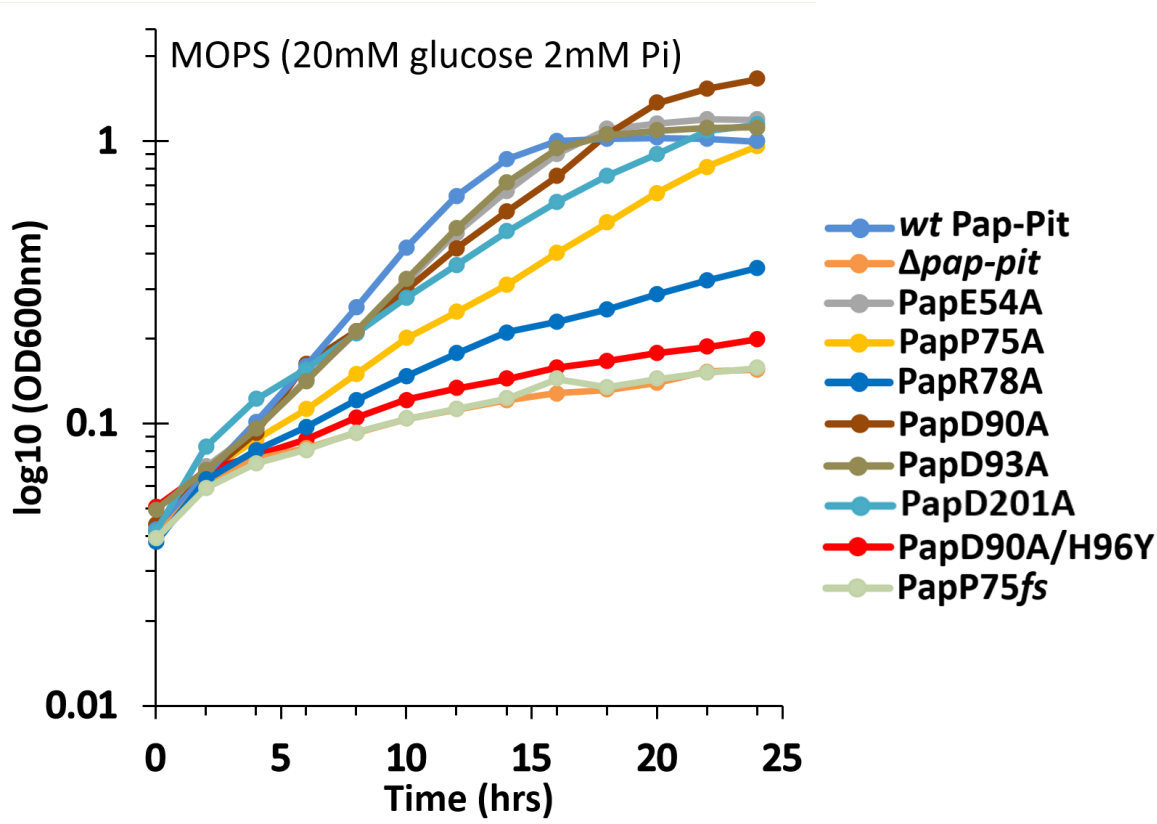


Table 3-21B

Strain	Velocity	StDev
RmP3926 (wildtype *pap-pit*)	3.12	0.142
RmP3926 (Δ*pap-pit*)	0.332	0.121
RmP3926 (PapE54A)	1.665	1.35
RmP3926 (PapT74A)	0.893	0.892
RmP3926 (PapP75A)	0.198	0.124
RmP3926 (PapR78A)	0.983	0.502
RmP3926 (PapD90A)	2.21	0.953
RmP3926 (PapD93A)	2.99	1.37
RmP3926 (PapD201A)	1.50	0.819
RmP3926 (PapD90A/H96Y)	0.422	0.388
RmP3926 (PapP75*fs*)	0.401	0.367

Figure 3-21. Phenotypic effects of Pap missense mutations on growth and Pi uptake of *S. meliloti*. A. RmP3926 (Δ*pap-pit*) with seven different Pap missense mutations were observed to have phenotypic effects on growth in MOPS-P2. **B.** Initial velocity (nmol Pi/mg protein/min) of Pi uptake was measured using radiolabeled Pi. Velocities reflected growth phenotypes except for RmP3926 (PapT74A) which showed poor uptake

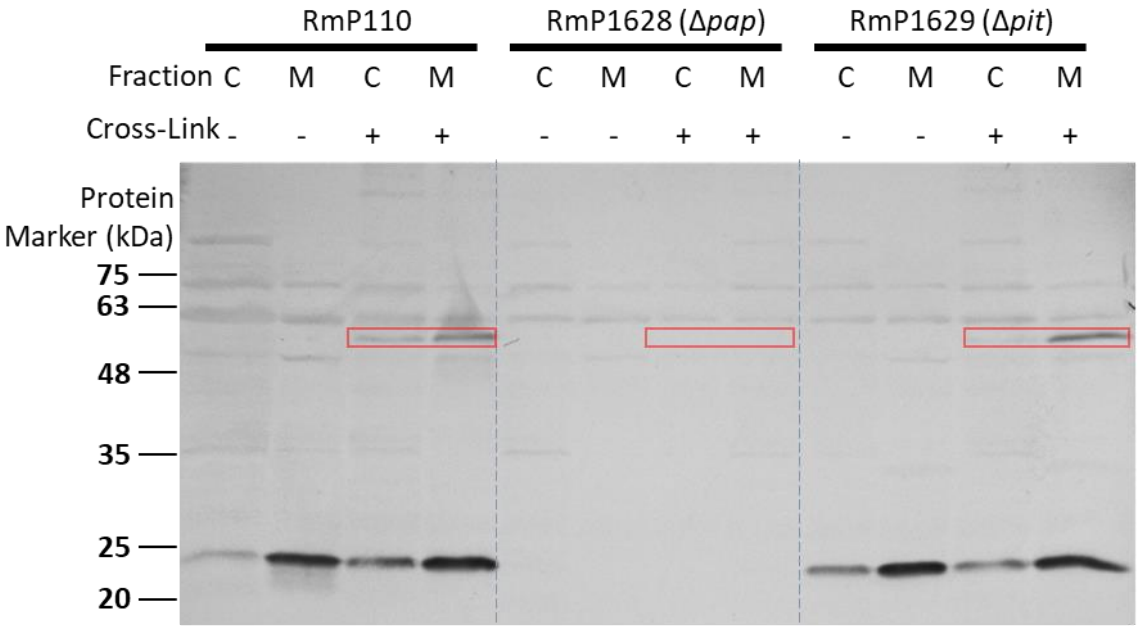


Figure 3-22. Pap proteins interactions by *in vivo* protein cross-linking with formaldehyde. Total protein from RmP110 (wt *pap-pit*), RmP1628 (Δ*pap*), and RmP1629 (Δ*pit*) were cross-linked with formaldehyde, fractionated by centrifugation into cytoplasmic (C) and membrane (M) fractions, then run on an SDS-PAGE prior to anti-Pap Western blot. A novel band at ~55kDa is present in the cross-linked fractions of RmP110 and RmP1629, but not RmP1628, indicating that this protein is likely SmPap cross-linked with another protein.

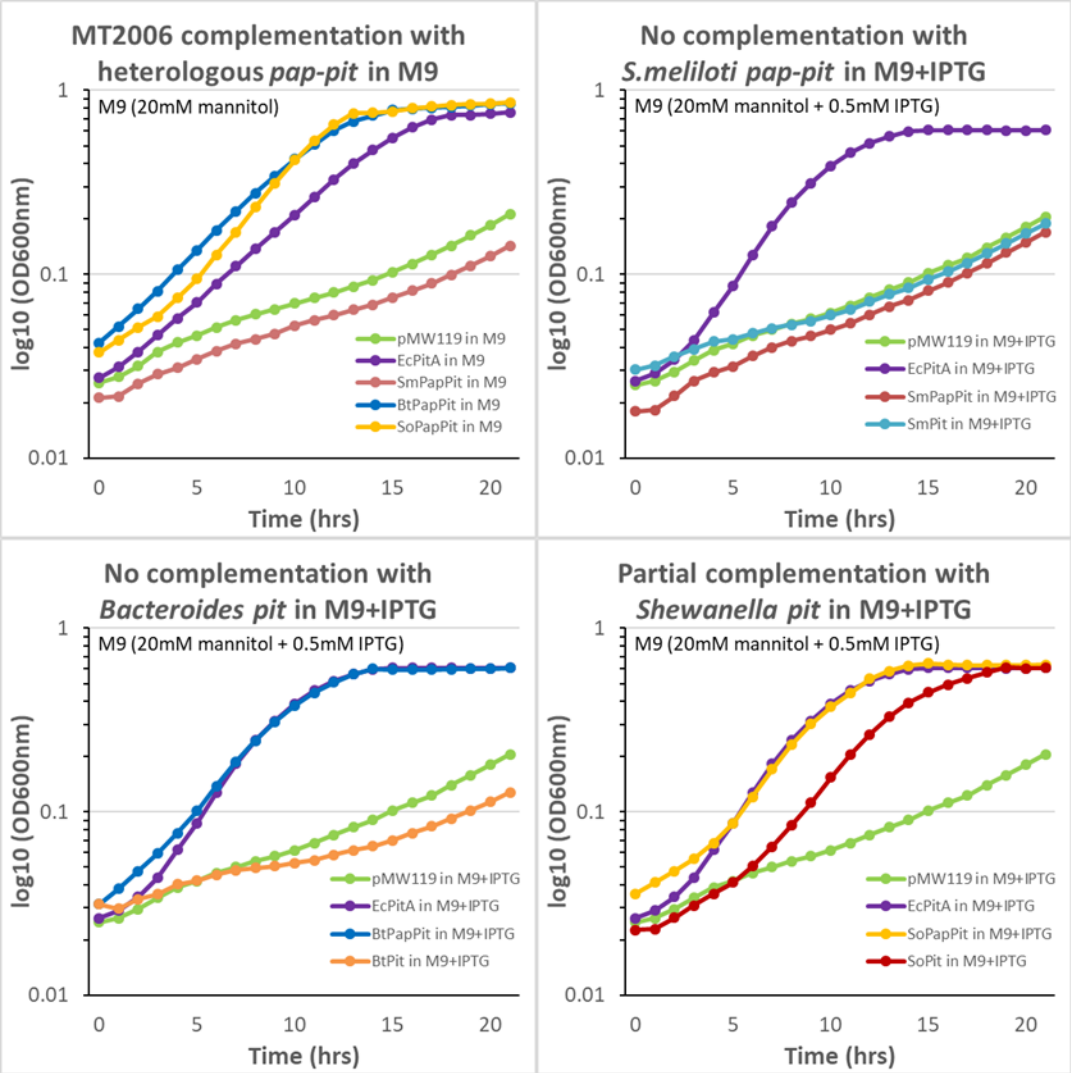
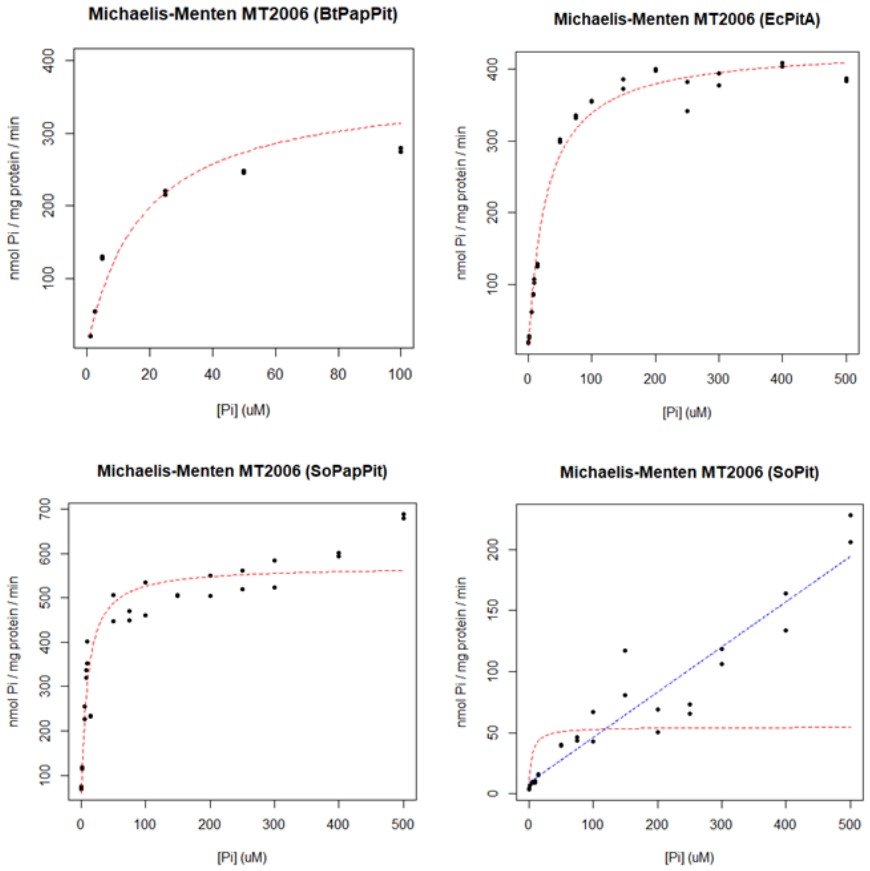


Figure 3-23. Complementation of Pi-uptake deficient *E. coli* MT2006 growth with heterologous Pap-Pit systems. BtPapPit, SoPapPit and EcPitA in pMW119 complemented the Pi uptake-deficient *E. coli* MT2006 when grown in M9. SmPapPit did not complement the growth of MT2006 even when grown in M9+IPTG. BtPit and SmPit also did not complement MT2006 when grown in M9+IPTG. SoPit partially complemented MT2006 when grown in M9+IPTG, with similar velocity but delayed stationary phase (10 hrs).

Table 3-24

Pi uptake system	Velocity (nmol Pi mg^{-1} min^{-1})	Standard Deviation
pMW119 (---)	0.274^{d}	0.037
SmPapPit	0.218^{d}	0.042
BtPapPit	18.9^{b}	0.924
SoPapPit	39.3^{a}	1.09
EcPitA	3.85^{c}	0.352
pMW119 (---) (+IPTG)	1.64^{c}	0.101
SmPit (+IPTG)	1.40^{c}	0.026
BtPit (+IPTG)	1.27^{c}	0.023
SoPit (+IPTG)	6.10^{b}	0.187
EcPitA (+IPTG)	77.5^{a}	10.8

Table 3-24. Initial Pi uptake velocities of heterologous Pap-Pit systems. Pi uptake of was only observed for MT2006 with BtPapPit, SoPapit or EcPitA, but not SmPapPit. Uptake was also only observed for SoPit when first inoculated in M9+IPTG, but not SmPit or BtPit. Initial Pi uptake velocity was determined in triplicate at 10µM. Statistical groupings are indicated by superscript letters determined by one-way ANOVA followed by a Tukey's post-hoc test ($p < 0.05$).



Strain	K_M (µM Pi)	V_{MAX} (nmol Pi mg protein^{-1} min^{-1})
MT2006 (BtPapPit)	17.2 ± 3.9***	368 ± 18***
MT2006 (SoPapPit)	8.5 ± 1.4***	571 ± 18***
MT2006 (SoPit)	3.9 ± 0.9	54.7 ± 2.9
MT2006 (EcPitA)	27.2 ± 2.1***	431 ± 7***

Figure 3-25. Michaelis-Menten kinetics of Pi transport via the heterologous Pit systems in *E. coli* MT2006. Kinetics of Pi uptake were measured in duplicate transport reactions with concentrations between 0.25-500µM Pi. BtPapPit and SoPapPit were found to have greater affinity to Pi compared to EcPitA. SoPit (inoculated in M9+IPTG) did not appear to become saturated. K_M and V_{MAX} were determined by non-linear regression of the Michaelis-Menten plot in R using the nls() function. *** $p < 0.001$

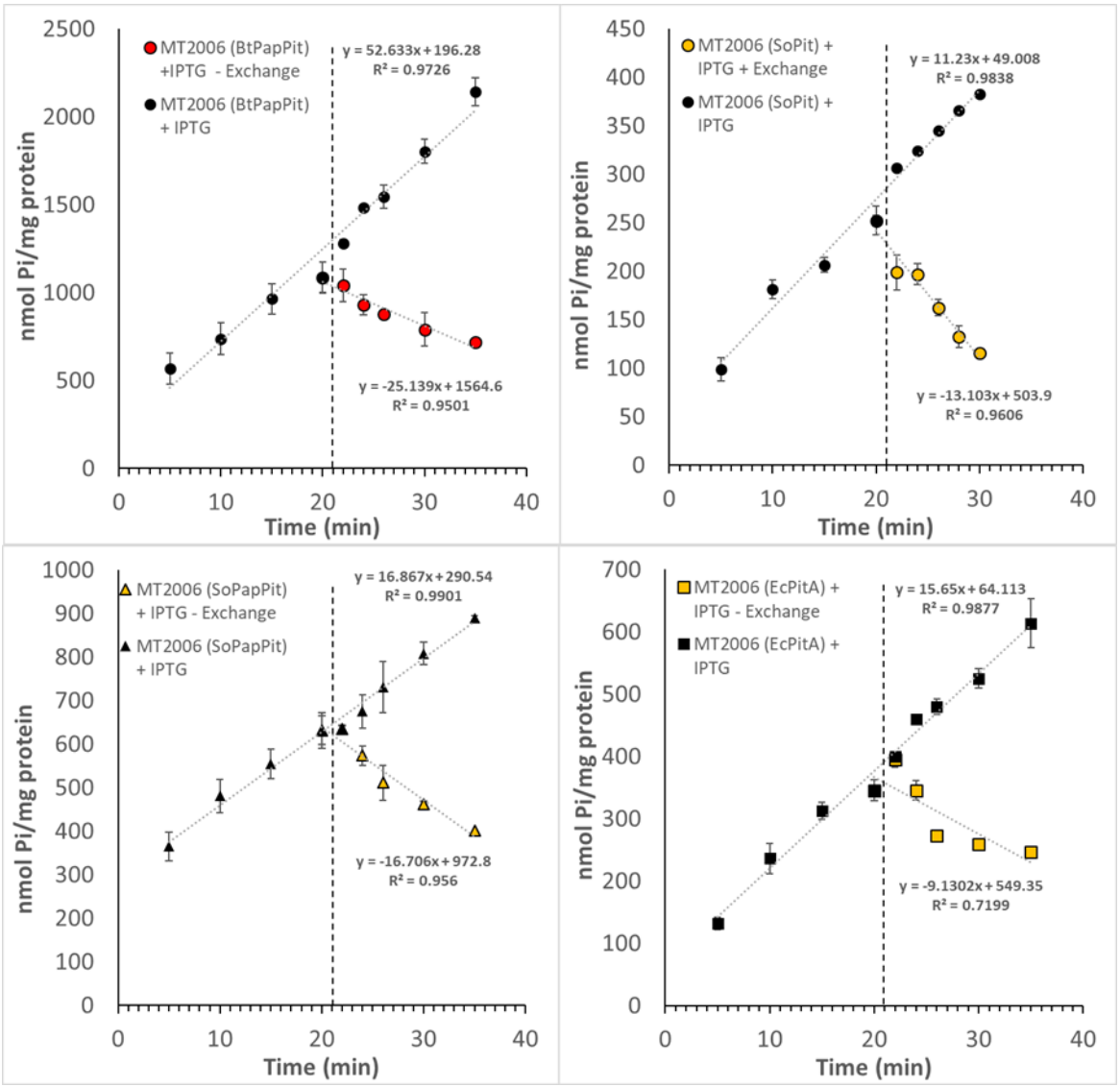


Figure 3-26. Exchange of Pi through the Pap-Pit systems. Accumulation and exchange of Pi was measured in transport reactions with 500µM Pi between 5 and 35 minutes in duplicate reactions. Exchange of Pi was induced by addition of 30-fold excess unlabeled Pi at 22minutes to one of two concurrent transport reactions. Exit of Pi was observed after addition of unlabeled Pi for BtPapPit, SoPapPit, SoPit and EcPitA, at an exit velocity similar to that of the uptake velocity at the same time points.

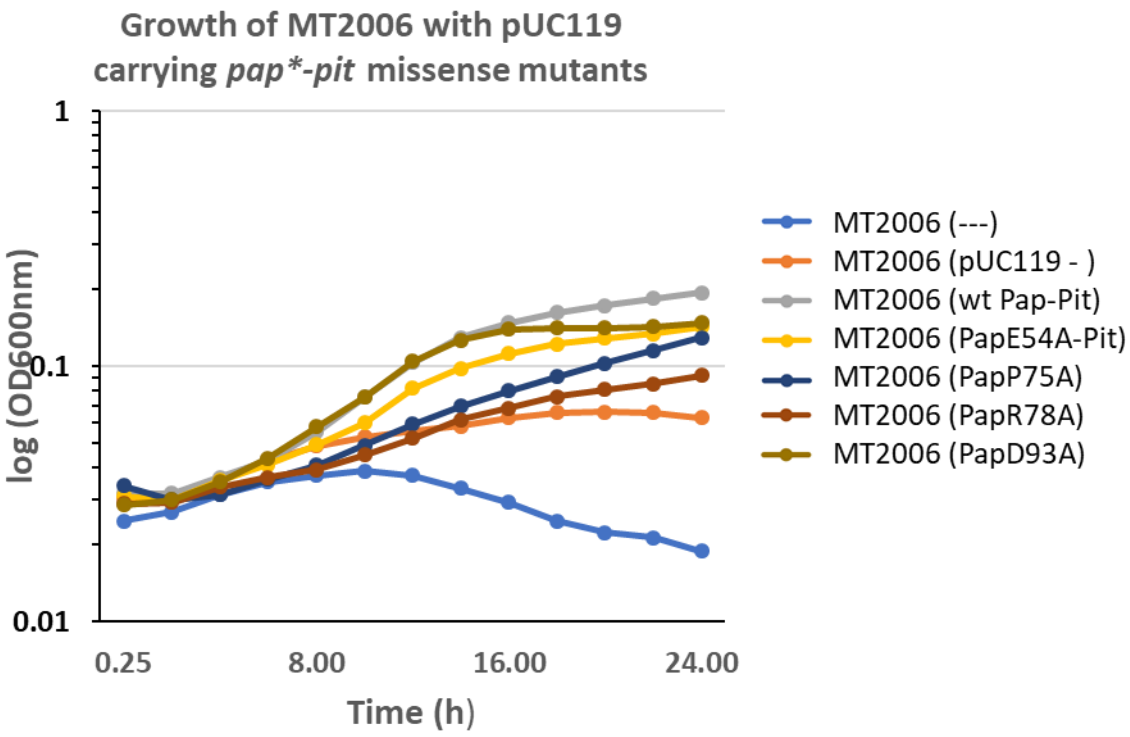


Figure 3-27. Phenotypic effects of *S. meliloti* Pap missense mutations expressed in *E. coli* MT2006. MT2006 with pUC119 carrying SmPap*Pit (with 14 different Pap missense mutations) were grown in M9. Only PapE54A, PapP75A, PapR78A, and PapD93A did not complemented MT2006 (pUC119) in M9, indicating that these mutations resulted in defective Pi uptake via SmPapPit.

Table 3-28

	Velocity	StDev
MT2006	0.548	0.183
MT2006 (pTH2826)	27.4	3.16
MT2006 (pUC119)	0.348	0.119
M2375 (PapE54A)	4.61	2.16
M2378 (PapD58A)	2.00	0.987
M2381 (PapT74A)	23.7	0.911
M2384 (PapP75A)	2.38	1.44
M2387 (PapR78A)	0.629	0.282
M2390 (PapD90A)	25.1	2.55
M2393 (PapD93A)	1.23	0.221
M2396 (PapD160A)	11.1	1.18
M2399 (PapE193A)	17.2	1.96
M2402 (PapD197A)	22.6	1.15
M2405 (PapD201A)	22.9	2.07
M2408 (PapD90A/H96Y)	7.86	0.590
M2411 (PapP75fs)	1.56	0.440

Table 3-28. Initial Pi uptake velocities of *S. meliloti* Pap missense mutations in *E. coli* MT2006. Initial Pi uptake velocities (nmol Pi/mg protein/min) of *E. coli* MT2006 carrying pUC119 with SmPapPit that had various Pap missense mutations. PapE54A, PapD58A, PapP75A, PapR78A, PapD93A and PapD90A/H96Y were observed to have reduced uptake velocity, compared to wildtype SmPapPit

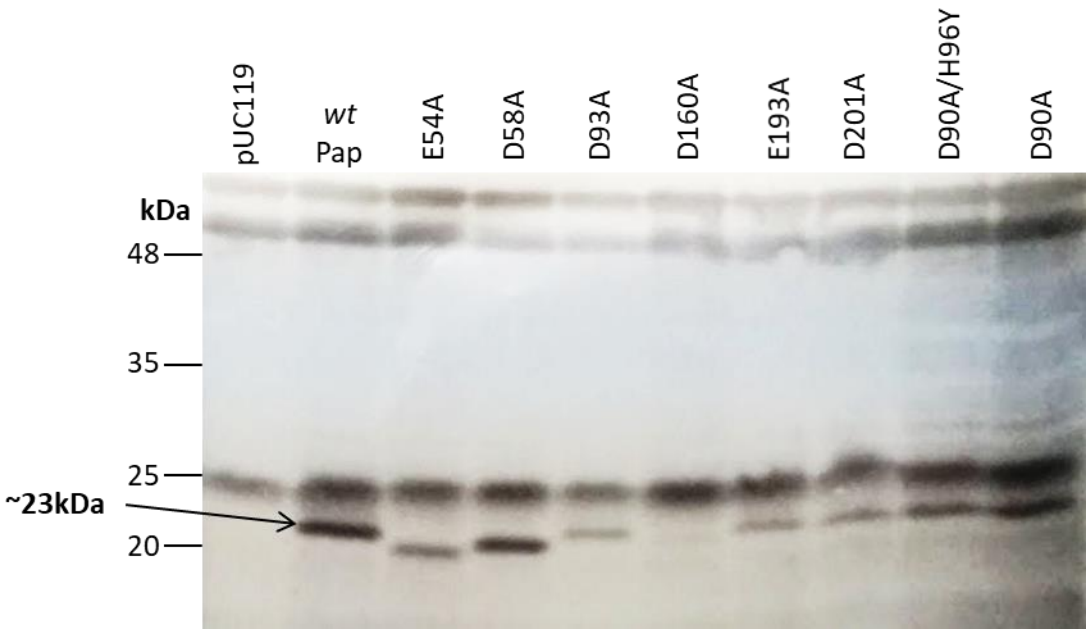


Figure 3-29. Anti-Pap Western blot of S. meliloti Pap missense mutants expressed in *E. coli* MT2006. Total protein from *E. coli* MT2006 with pUC119 carrying SmPap*Pit (with 14 different Pap missense mutations) were run on an SDS-PAGE followed by anti-Pap western blot. Compared to MT2006 (SmPapPit) some mutants (PapE54A, D58A, D160A) all had a Pap protein which was smaller in size at 20kDa rather than 23kDa.

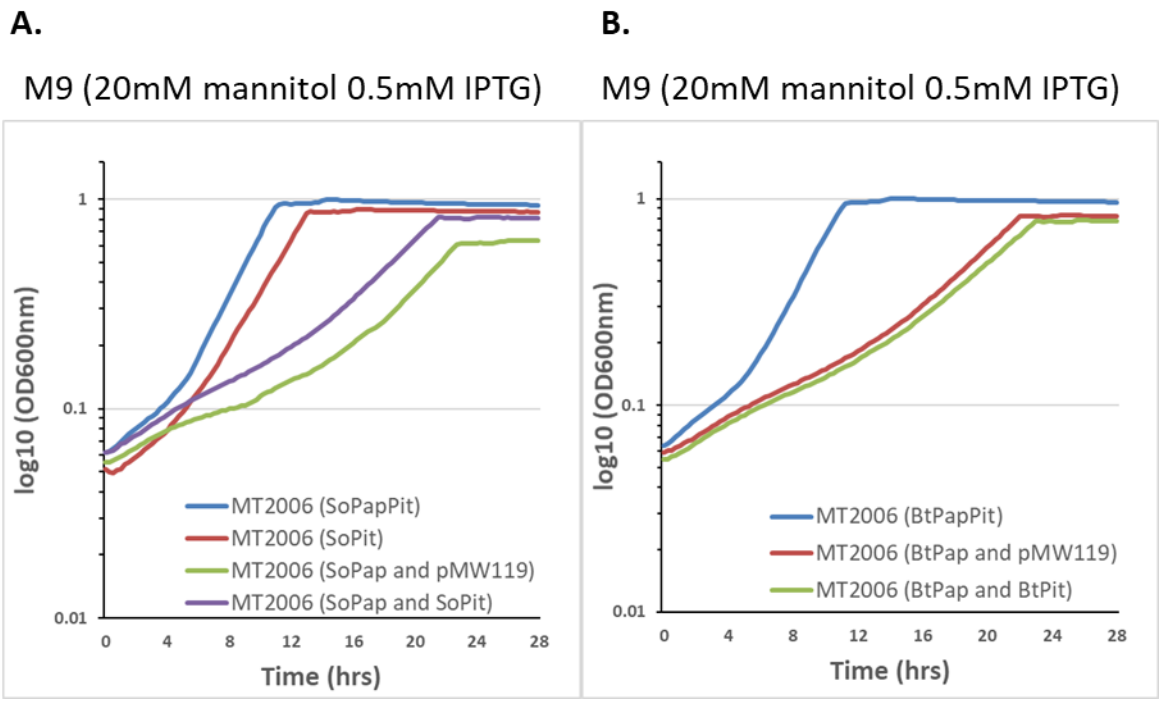


Figure 3-30. Growth complementation of *E. coli* MT2006 co-transformed with cognate heterologous Pap and Pit genes. Growth experiments in M9-mannitol with IPTG of MT2006 with *pap* and *pit* genes encoded on separated plasmids. **A.** *E. coli* MT2006 with SoPapPit and SoPit were observed to grow, however MT2006 (SoPap and SoPit) had a growth curve similar to *E. coli* MT2006 (SoPap and pMW119). This is strange since SoPit is present, but does not grow as well as MT2006 (SoPit) alone. **B.** Additionally, no complementation was observed for *E. coli* MT2006 (BtPap and BtPit).

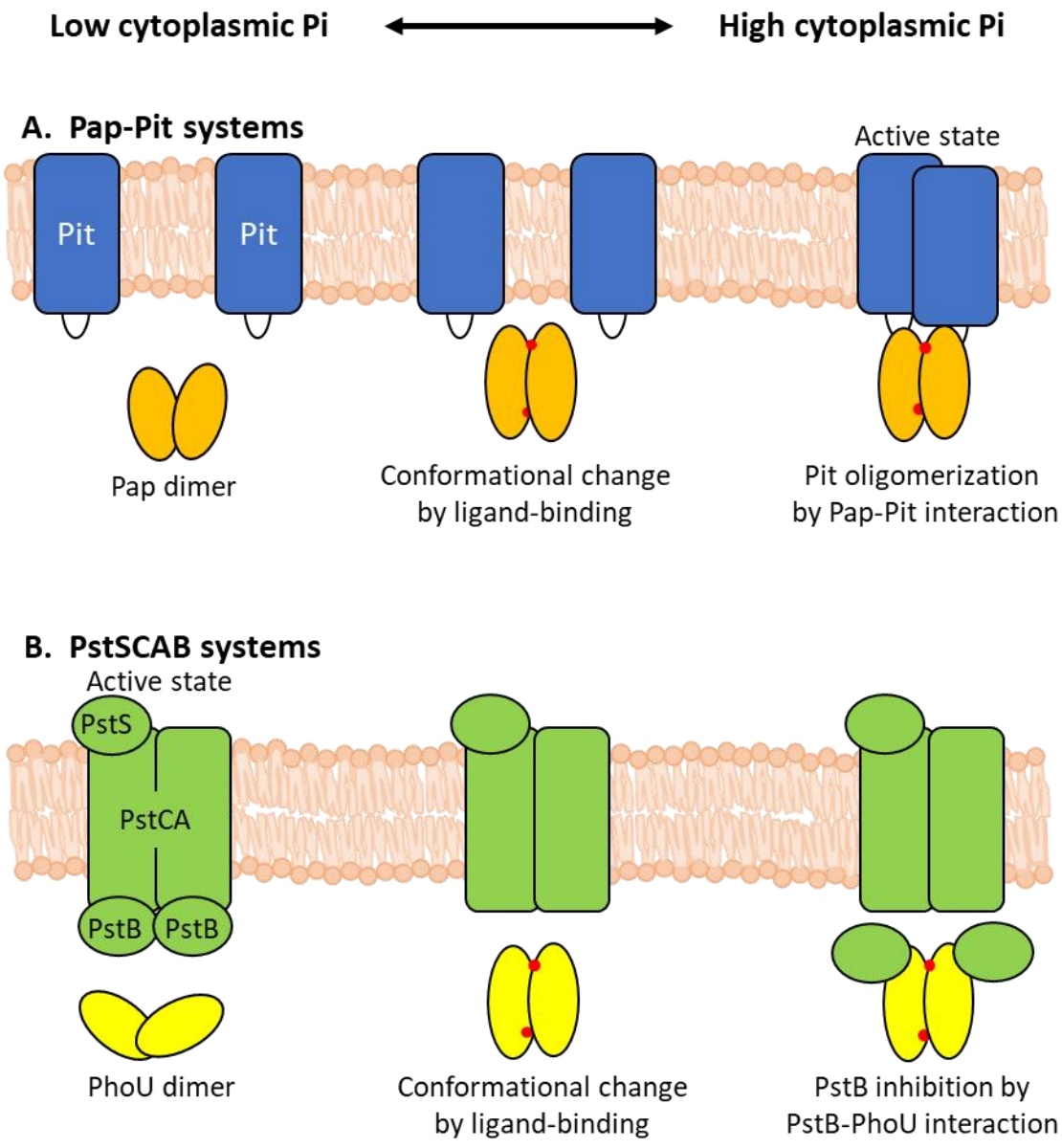


Figure 4-1. Model of the mechanism of action for Pap and PhoU-like proteins. The model of PhoU-like proteins involves sensing high cytoplasmic Pi by binding to an phosphate-metabolite that results in dimerization. **A.** For Pap-Pit, ligand-binding of Pap and an interaction with Pit results in dimerization/activation of the transporter. This transport system is also inactive when cytoplasmic Pi is low. **B.** For PstSCAB-PhoU, ligand-binding of PhoU and an interaction with PstB when cytoplasmic Pi is high, results in inhibition of PstB and preventing additional Pi uptake.

Appendix

Appendix Table 1 – Comprehensive list of Primers used in this study.

PitR-PstI	AATTCCTGCAGGGCGACGAGGTCCG
XbaI-PitF	CCGCTCTAGAATGGATGCGACGCTCG
phoA-lacZseq	GACCAGACCGATCTCTTCTACACC
SmPit135	AATTCCTGCAGCAGCAGGCCTTGC
SmPit193	AATTCCTGCAGGTTGCCGCCATGG
SmPit238	AATTCCTGCAGCTTCGAGCCCATCG
pitHis-upF	CCGGCCTGGCCAAGAC
pitHis-dnR	CCAGTGCCAAGCTTCTTTCG
pitHis-dnF	GGGGCGGGGCACCACCATCACCACCACCACCACCATCACTGACGCTTTCGCCG GGC
pitHis-upR	TCAGTGATGGTGGTGGTGGTGGTGATGGTGGTGCCCCGCCCCGGCGACGAGGT CCGC
pitscreen-dnR	TGTCGGGCGGCATGGCG
pitscreen-dnF	TCGTCGCCGAGGCCATTCC
papscreen-upR	GAAGGCGAGCGTCGCATCC
papscreen-upF	ATTTTCGCGCAGGGCGCTGC
pit-pJP2R	TGCATCGATAGATCTCGTCAGGCGACGAGG
pit-pJP2F	GGATGGATCGCTGATGGATGCGACGCTCG
pap-promR	AGCGTCGCATCCATCAGCGATCCATCC
pap-promF	CATGGATCCAAGCTTCCGTTCTCGCCGG
pit-pJP2.2R	TGGCATGCATCGATAGATCTCGTCAGGCGACGAGG
pit-pJP2.2F	GACAGTTGGATGGATCGCTGATGGATGCGACGCTCG
orfAprom-R	GAAGGCGAGCGTCGCATCCATCAGCGATCCATCC
orfAprom-F	GGTACCATGGATCCAAGCTTCCGTTCTCGCCGG
pitHis-dnRnew	ACGTTGTAAAACGACGGCCAGTGCCAAGC
pitHis-upFnew	TCGGCGCCGGCCTGGCCAAGACC
SP002R	GATGCCGCAGGCGATTACAT
SP002F	TTTGATTTCACGGGTTGGGG
pitscreenR-3.0	CGCTGCCACGATTGC
papscreenF-3.0	GCGCTGCATGTCACC
pitscreen-R2.0	CGCTGCCACGATTGC
papscreenF-2.0	TCGGTCGTCATGTCC
pitHis-BglII	TGGCATGCATCGATAGATCTGCGTCAGTGATGG
SmpromF	TATGACCATGATTACGCCAAGCTGGTCAGCACCACATGGCG
SmpromR	TGCCGTTTACCTATTAAGCTCAGCGATCCATCCAACTGTCG
SmPromR-Bt	ACATTAAATTGTCGCAAGCTTCAGCGATCCATCCAACTGTCG
BtPitR	ATCCCGGGAACTTGAAGCAGAATTTATTTCAGG
BtpapF	AGCAAGCTTGCGACAATTTAATGTACTCATATTTATG
PromF	AGCTATGACCATGATTACGAATTCGGTCAGCACCACATGG
XmaI-SmPapF	TACCCGGGGATGGATGGATCGCTGATGC
Rev-Sm	AGGTCGACTCTAGAGGATCCTGCCCGGCGAAAGCG
Rev-So	AGGTCGACTCTAGAGGATCCCGGGCCAGTTTTGTGC
SmPitR-HindIII	CCAAGCTTGCGAAAGCGTCAGGCGACG
Rev-Bt	CAGGTCGACTCTAGAGGATCCATTGAAGCAGAATTTATTTCAGG
pitdnR-pPLE01	GTCGACGGTATCGATAAGCTTCTTTCGACGTCGATCTCG
pitdnF-pPLE01	AATGGCGAATGGTGAAAGCTTCGCTTTCGCCGGGCAC
BtPitR-pMW	ATCCCGGGAACCAGAATTTATTTCAGGAACAGAG

BtPitF-pMW	AGCAAGCTTGCGGCAGAGCAAGACACATGG
pitR-HindIII	CCAAGCTTTCAGGCGACGAGGTC
pitTruncR-HindIII	CCAAGCTTCAGGTAACGATGTTCCAGG
papR-HindIII	CCAAGCTTTCAGACGTTCTCGATCAC
XbaI-papTruncF	CCTCTAGATTCGATCCCGGCATGC
XbaI-promF	CCTCTAGACCGTTCTCGCCGGTC
Strep_coding	GGAGAACCTGTATTTCCAGTGGAGCCACCCGCAGTTCGAGAAGTGAA
FLAG_coding	GGAGAACCTGTATTTCCAGGATTATAAGGACGATGACGATAAGTGAA
1D4_coding	GGAGAACCTGTATTTCCAGACCGAGACGAGCCAGGTCGCCCCGGCCTGAA
FLAG_complement	AGCTTTCACTTATCGTCATCGTCCTTATAATCCTGGAAATACAGGTTCTCCTGCA
1D4_complement	AGCTTTCAGGCCGGGGCGACCTGGCTCGTCTCGGTCTGGAAATACAGGTTCTCCTGCA
Strep_complement	AGCTTTCACTTCTCGAACTGCGGGTGGCTCCACTGGAAATACAGGTTCTCCTGCA
pTrcSC-seqfor	GAGCTGTTGACAATTAATCATCC
pTrcSC-seqrev	ATCAGACCGCTTCTGCGTTC
pitAFor-pMW119	ACCATGATTACGCCAAGCTGAGCGTCCTATGCTAC
pitARev-pMW119	ATTCGAGCTCGGTACCCGGTTACAGGAACTGCAAGG
PitSeq-F	GCGCGTATCGGCGGTGC
papI209I-F	GCAAACGAAATAAGCGGC
papI209STOP-F	GCAAACGAATAGAGCGGC
papI209I-R	GCCGCTTATTTCGTTTGC
papI209STOP-R	GCCGCTCTATTCGTTTGC
pcaprom-R	CCGGGTACCGAGCTCGAATTGGCAGTTCTCCTCCG
pcaprom-F	TATGACCATGATTACGAATTGGATCGTATAACCTCCTG
BamHI-pcapromR	AGAGGATCCGGCAGTTCTCCTCCG
EcoRI-pcapromF	ACGAATTCGGATCGTATAACCTCCTG
EcoRV-PlacF	TACTAGTATTTAAATGATGCAGTGAGCGCAACGC
BtPapR-EcoRV	TTAATTAACCGATCGGATCTTTTATGAATACTTTACGATCAGG
SoPapR-EcoRV	TTAATTAACCGATCGGATTATTTATTAAACGCGAGCCAGC
papF-XbaI	TCCTCTAGATTGGATGGATCGCTGATG
pitF-XbaI	TCCTCTAGAATCGAGAACGTCTGATGG
SmSp_pTH1937-F	GTCCGGTGCCCTGAATGAACGCGAAGCCCGTTTAAACC
FLAG_complement	AGCTTTCACTTATCGTCATCGTCCTTATAATCCTGGAAATACAGGTTCTCCTGCA
1D4_complement	AGCTTTCAGGCCGGGGCGACCTGGCTCGTCTCGGTCTGGAAATACAGGTTCTCCTGCA
Strep_complement	AGCTTTCACTTCTCGAACTGCGGGTGGCTCCACTGGAAATACAGGTTCTCCTGCA
pTrcSC-seqfor	GAGCTGTTGACAATTAATCATCC
pTrcSC-seqrev	ATCAGACCGCTTCTGCGTTC
pitAFor-pMW119	ACCATGATTACGCCAAGCTGAGCGTCCTATGCTAC
pitARev-pMW119	ATTCGAGCTCGGTACCCGGTTACAGGAACTGCAAGG
PitSeq-F	GCGCGTATCGGCGGTGC
papI209I-F	GCAAACGAAATAAGCGGC
papI209STOP-F	GCAAACGAATAGAGCGGC
papI209I-R	GCCGCTTATTTCGTTTGC
papI209STOP-R	GCCGCTCTATTCGTTTGC
pcaprom-R	CCGGGTACCGAGCTCGAATTGGCAGTTCTCCTCCG
pcaprom-F	TATGACCATGATTACGAATTGGATCGTATAACCTCCTG

BamHI-pcapromR	AGAGGATCCGGCAGTTCTCCTCCG
EcoRI-pcapromF	ACGAATTCGGATCGTATAACCTCCTG
EcoRV-PlacF	TACTAGTATTTAAATGATGCAGTGAGCGCAACGC
BtPapR-EcoRV	TTAATTAACCGATCGGATCTTTTATGAATACTTTACGATCAGG
SoPapR-EcoRV	TTAATTAACCGATCGGATTATTTATTAAACGCGAGCCAGC
papF-XbaI	TCCTCTAGATTGGATGGATCGCTGATG
pitF-XbaI	TCCTCTAGAATCGAGAACGTCTGATGG
SmSp_pTH1937-F	GTCCGGTGCCCTGAATGAACGCGAAGCCCGTTTAAACC
SmSp_pTH1937-R	ATAGCCGCGCTGCCTCGTCCATCCGGTGATTGATTGAGC
ScPitH1R-XmaI	ATCCCGGGAACCTACAGGACCGCCAGG
HindIII-ScPitF	AGCAAGCTTGCGCGCCGTCAAGGAGTCC
HindIII-ScPapF	AGCAAGCTTGCGCATGCGCTTTCGTCTGAC
SpR-R-XmaI	GCATCCCGGGGATCCGGTGATTGATTGAG
pitARev-pMW119	ATTCGAGCTCGGTACCCGGTTACAGGAACTGCAAGG
PitSeq-F	GCGCGTATCGGCGGTGC
papI209I-F	GCAAACGAAATAAGCGGC
papI209STOP-F	GCAAACGAATAGAGCGGC
papI209I-R	GCCGCTTATTTCGTTTGC
papI209STOP-R	GCCGCTCTATTCGTTTGC
pcaprom-R	CCGGGTACCGAGCTCGAATTGGCAGTTCTCCTCCG
pcaprom-F	TATGACCATGATTACGAATTGGATCGTATAACCTCCTG
BamHI-pcapromR	AGAGGATCCGGCAGTTCTCCTCCG
EcoRI-pcapromF	ACGAATTCGGATCGTATAACCTCCTG
EcoRV-PlacF	TACTAGTATTTAAATGATGCAGTGAGCGCAACGC
BtPapR-EcoRV	TTAATTAACCGATCGGATCTTTTATGAATACTTTACGATCAGG
SoPapR-EcoRV	TTAATTAACCGATCGGATTATTTATTAAACGCGAGCCAGC
papF-XbaI	TCCTCTAGATTGGATGGATCGCTGATG
pitF-XbaI	TCCTCTAGAATCGAGAACGTCTGATGG
SmSp_pTH1937-F	GTCCGGTGCCCTGAATGAACGCGAAGCCCGTTTAAACC
SmSp_pTH1937-R	ATAGCCGCGCTGCCTCGTCCATCCGGTGATTGATTGAGC
ScPitH1R-XmaI	ATCCCGGGAACCTACAGGACCGCCAGG
HindIII-ScPitF	AGCAAGCTTGCGCGCCGTCAAGGAGTCC
HindIII-ScPapF	AGCAAGCTTGCGCATGCGCTTTCGTCTGAC
SpR-R-XmaI	GCATCCCGGGGATCCGGTGATTGATTGAG
XhoI-SpR-F	GCATCTCGAGGCGAAGCCCGTTTAAACC
XmaI-SpR-F	GCATCCCGGGGCGAAGCCCGTTTAAACC
XmaI-SpR-R	GCATCCCGGGATCCGGTGATTGATTGAGC
PitR	GAA ATA GCA GAG CGC CGA G
PitF	CGC TTT TCT TCG ACT TCC TC
PIT500-dnR	GTA AGC TTC TTT CGA CGT CGA TCT CG
PapF	CTG TTT CGC AAG CTC CTC C
500PAP500_R	GTA AGC TTA CGA TCA GCA CCA ACA GC
PAP500-upF	GTT CTA GAC CGA TCT CCA CCG ACC
PapEnd-F	GAG ATC TAC GGC GAA CTG
PhoA-R	CTA AGA GAA TCA CGC AGA GC

Appendix Table 3 – Full list of Pap-Pit orthologues identified

Organisms	pap	pit
Alicyclobacillus acidocaldarius subsp. acidocaldarius DSM 446	Aaci_0054	Aaci_0053
Alicyclobacillus acidocaldarius subsp. acidocaldarius DSM 446	Aaci_2627	Aaci_2626
Glutamicibacter arilaitensis Re117	AARI_01240	AARI_01230
Aggregatibacter aphrophilus NJ8700	NT05HA_0317	NT05HA_0316
Aggregatibacter actinomycetemcomitans D11S-1	D11S_1778	D11S_1779
Paenarthrobacter aurescens	AAur_0200	AAur_0199
Acidovorax citrulli	Aave_1880	Aave_1881
Candidatus Koribacter versatilis	Acid345_3130	Acid345_3131
Acidobacteria bacterium DSM 100886	LuPra_03057	LuPra_03058
Aciduliprofundum boonei	Aboo_0036	Aboo_0037
Aciduliprofundum boonei	Aboo_1146	Aboo_1147
Alcanivorax borkumensis	ABO_2304	ABO_2305
Azospirillum brasilense Sp245	AZOBR_p150029	AZOBR_p150030
Acidothermus cellulolyticus	Acel_1535	Acel_1534
Acidothermus cellulolyticus	Acel_2028	Acel_2027
Aciduliprofundum sp. MAR08-339	AciM339_0038	AciM339_0039
Aciduliprofundum sp. MAR08-339	AciM339_1473	AciM339_1474
Pseudarthrobacter chlorophenolicus	Achl_0368	Achl_0367
Azotobacter chroococcum	Achr_36220	Achr_36210
Anaeromyxobacter dehalogenans 2CP-C	Adeh_4227	Adeh_4228
Alcanivorax dieselolei	B5T_00640	B5T_00639
Alicycliphilus denitrificans BC	Alide_1199	Alide_1200
Alkalilimnicola ehrlichii	Mlg_2827	Mlg_2826
Adlercreutzia equolifaciens	AEQU_1521	AEQU_1522
Aeromicrobium erythreum	AERYTH_14370	AERYTH_14365
Actinobacillus equuli	ACEE_03460	ACEE_03455
Actinoplanes friuliensis	AFR_40800	AFR_40795
Archaeoglobus fulgidus DSM 4304	AF_1799	AF_1798
Anaeromyxobacter sp. Fw109-5	Anae109_4377	Anae109_4378
Archangium gephyra	AA314_09218	AA314_09219
Archangium gephyra	AA314_09218	AA314_09220
Aeromonas hydrophila subsp. hydrophila ATCC 7966	AHA_3771	AHA_3772
Arcanobacterium haemolyticum	Arch_0220	Arch_0219
Acidianus hospitalis	Ahos_2112	Ahos_2113
Azospirillum lipoferum	AZOLI_p40213	AZOLI_p40212
Alloactinosynnema sp. L-07	CRK57760	CRK57761
Alcanivorax sp. NBRC	AS19_23670	AS19_23680
Allochromatium vinosum	Alvin_1247	Alvin_1246
Amycolatopsis mediterranei U32	AMED_4281	AMED_4282
Aeromonas media	B224_0256	B224_0255

Actinoplanes missouriensis	AMIS_76830	AMIS_76840
Actinomyces meyeri	ADJ76_08665	ADJ76_08660
Anaeromyxobacter sp. K	AnaeK_4360	AnaeK_4361
Aneurinibacillus sp. XH2	ACH33_13160	ACH33_13165
Bradyrhizobium oligotrophicum	S58_42270	S58_42260
Alcanivorax pacificus	S7S_02230	S7S_02225
Alpha proteobacterium HIMB59	HIMB59_00004450	HIMB59_00004460
Actinobacillus pleuropneumoniae L20 (serotype 5b)	APL_0718	APL_0719
Pseudarthrobacter phenanthrenivorans	Asphe3_03550	Asphe3_03540
Archaeoglobus profundus	Arcpr_1237	Arcpr_1238
Archaeoglobus profundus	Arcpr_1272	Arcpr_1271
Accumulibacter phosphatis	CAP2UW1_2086	CAP2UW1_2085
Accumulibacter phosphatis	CAP2UW1_3732	CAP2UW1_3733
Accumulibacter phosphatis	CAP2UW1_3786	CAP2UW1_3785
Accumulibacter phosphatis	CAP2UW1_3787	CAP2UW1_3788
Accumulibacter phosphatis	CAP2UW1_3789	CAP2UW1_3788
Candidatus Azobacteroides pseudotrichonymphae (Coptotermes formosanus)	CFPG_236	CFPG_237
Atopobium parvulum	Apar_0042	Apar_0043
Atopobium parvulum	Apar_0386	Apar_0385
Agrobacterium radiobacter	Arad_9331	Arad_9333
Arachidicoccus sp. BS20	A9P82_05200	A9P82_05205
Actinomyces radicidentis	AXF14_06550	AXF14_06555
Arthrobacter sp. PAMC 25486	ART_2529	ART_2530
Arthrospira platensis	NIES39_A04380	NIES39_A04370
Arthrobacter sp. Rue61a	ARUE_c01940	ARUE_c01930
Arthrobacter sp. FB24	Arth_0207	Arth_0206
Aeromonas salmonicida	ASA_0518	ASA_0517
Actinoplanes sp. SE50/110	ACPL_7964	ACPL_7965
Actinotignum schaalii	FB03_05670	FB03_05675
Actinobacillus suis H91-0380	ASU2_03490	ASU2_03485
Aeromonas schubertii	WL1483_4623	WL1483_4708
Archaeoglobus sulfaticallidus	Asulf_00411	Asulf_00410
Archaeoglobus sulfaticallidus	Asulf_01051	Asulf_01050
Actinobacillus succinogenes	Asuc_1078	Asuc_1077
Caldicellulosiruptor bescii	Athe_1484	Athe_1485
Agrobacterium tumefaciens Ach5	Ach5_43460	Ach5_43450
Azospirillum thiophilum	AL072_24680	AL072_24685
Agrobacterium fabrum	Atu4634	Atu4633
Archaeoglobus veneficus	Arcve_1337	Arcve_1336
Aeromonas veronii	B565_3572	B565_3571
Aliivibrio wodanis	AWOD_I_0361	AWOD_I_0360

Acetobacterium woodii	Awo_c07440	Awo_c07430
Azospirillum sp. B510	AZL_d01070	AZL_d01080
Bacteroidales bacterium CF	BRDCF_p1097	BRDCF_p1098
Bacillus atrophaeus 1942	BATR1942_04100	BATR1942_04095
Burkholderia ambifaria AMMD	Bamb_1806	Bamb_1805
Bacillus anthracis Ames	BA_1472	BA_1471
Bacillus amyloliquefaciens DSM 7	BAMF_1373	BAMF_1372
Bathyarchaeota archaeon BA1	AOA65_1858	AOA65_1856
Bacillus velezensis FZB42	RBAM_012700	RBAM_012690
Brevibacillus brevis	BBR47_36630	BBR47_36620
Brevibacillus brevis	BBR47_49120	BBR47_49130
Bifidobacterium bifidum S17	BBIF_0800	BBIF_0799
Bifidobacterium breve ACS-071-V-Sch8b	HMPREF9228_1001	HMPREF9228_1002
Paraburkholderia caribensis	K788_0002564	K788_0002565
Bacillus cereus ATCC 14579	BC1454	BC1453
Bacteroides cellulosilyticus	BcellWH2_01256	BcellWH2_01255
Bacillus coagulans 2-6	BCO26_2267	BCO26_2268
Burkholderia cenocepacia AU1054	Bcen_6210	Bcen_6211
Burkholderia contaminans	NL30_01010	NL30_01005
Burkholderia contaminans	NL30_13490	NL30_13495
Burkholderia cepacia GG4	GEM_1556	GEM_1557
Beutenbergia cavernae	Bcav_3420	Bcav_3421
Bacillus cytotoxicus	Bcer98_1179	Bcer98_1178
Burkholderia dolosa	AK34_1256	AK34_1257
Bacteroides dorei HS1_L_1_B_010	EL88_06790	EL88_06795
Bacillus endophyticus	BEH_04145	BEH_04150
Brachybacterium faecium	Bfae_07400	Bfae_07390
Paraburkholderia fungorum	OI25_3339	OI25_3340
Burkholderia gladioli BSR3	bgla_1g23420	bgla_1g23410
Burkholderia glumae BGR1	bglu_1g13990	bglu_1g14000
Bacteroides helcogenes	Bache_2734	Bache_2733
Bradyrhizobium icense	LMTR13_22905	LMTR13_22900
Bacillus infantis	N288_06175	N288_06170
Bradyrhizobium diazoefficiens USDA 110	blr4042	blr4043
Bradyrhizobium japonicum USDA 6	BJ6T_57760	BJ6T_57750
Blastomonas sp. RAC04	BSY18_2550	BSY18_2551
Bacillus paralicheniformis	BaLi_c15040	BaLi_c15030
Bacillus licheniformis ATCC 14580	BL03744	BL03745
Bifidobacterium longum NCC2705	BL0904	BL0905
Brevibacillus laterosporus	BRLA_c025750	BRLA_c025740
Brevibacillus laterosporus	BRLA_c033660	BRLA_c033650

Brevibacterium linens	A2T55_15235	A2T55_15230
Burkholderia mallei ATCC 23344	BMA1399	BMA1398
Bacillus methanolicus	BMMGA3_03750	BMMGA3_03745
Bacillus megaterium QM B1551	BMQ_4710	BMQ_4709
Burkholderia multivorans ATCC 17616 (JGI)	Bmul_1405	Bmul_1406
Bacillus mycoides 219298	DJ92_4102	DJ92_4101
Brevundimonas naejangsanensis	DA69_07660	DA69_07655
Bacteroides ovatus	Bovatus_04842	Bovatus_04841
Burkholderia oklahomensis E0147	DM82_1554	DM82_1553
Bacillus oceanisediminis	A361_06485	A361_06480
Paraburkholderia phymatum	Bphy_0916	Bphy_0917
Beta proteobacterium CB	D521_1197	D521_1196
Burkholderia pseudomallei K96243	BPSL1463	BPSL1464
Bacillus pumilus SAFR-032	BPUM_1162	BPUM_1161
Paraburkholderia phenoliruptrix	BUPH_02318	BUPH_02317
Paraburkholderia phytofirmans	Bphyt_1791	Bphyt_1792
Burkholderia pyrrocinia	ABD05_15500	ABD05_15495
Paraburkholderia rhizoxinica	RBRH_03690	RBRH_03689
Bifidobacterium scardovii	BBSC_2525	BBSC_2526
Bacillus simplex	UP17_04635	UP17_04630
Bacillus simplex	UP17_04635	UP17_04640
Bacillus smithii	BSM4216_0930	BSM4216_0929
Bacillus subtilis subsp. subtilis 168	BSU12850	BSU12840
Burkholderia thailandensis E264	BTH_I2184	BTH_I2185
Bacteroides thetaiotaomicron	BT_4638	BT_4637
Bacillus thuringiensis 97-27	BT9727_1336	BT9727_1335
Bibersteinia trehalosi USDA-ARS-USMARC-192	WQG_9710	WQG_9700
Kyrpidia tusciae	Btus_2360	Btus_2359
Bacillus toyonensis	Btoyo_4094	Btoyo_4093
Burkholderia lata	Bcep18194_A5169	Bcep18194_A5168
Burkholderia lata	Bcep18194_B0593	Bcep18194_B0592
Burkholderia vietnamiensis G4	Bcep1808_1795	Bcep1808_1794
Blastochloris viridis	BVIR_2165	BVIR_2166
Barnesiella viscericola	BARVI_11305	BARVI_11310
Bacteroides vulgatus	BVU_2705	BVU_2706
Bacillus weihenstephanensis KBAB4	BcerKBAB4_1377	BcerKBAB4_1376
Paraburkholderia xenovorans LB400	Bxe_A2396	Bxe_A2395
Chlamydia abortus S26/3	CAB060	CAB061
Clostridium acetobutylicum ATCC 824	CA_C3094	CA_C3093
Clostridium autoethanogenum	CAETHG_3326	CAETHG_3327

Catenulispora acidiphila	Caci_0739	Caci_0740
Caulobacter sp. K31	Caul_1415	Caul_1414
Cyanobacterium aponinum	Cyan10605_2537	Cyan10605_2538
Collimonas arenae	LT85_2137	LT85_2138
Chlamydia avium	M832_08610	M832_08620
Clostridium beijerinckii NCIMB 8052	Cbei_1774	Cbei_1775
Coxiella burnetii RSA 493	CBU_0015	CBU_0014
Clostridium butyricum	ATN24_01125	ATN24_01120
Clostridium baratii	U729_182	U729_181
Cupriavidus basilensis	RR42_m2546	RR42_m2545
Symbiobacter mobilis	Cenrod_1807	Cenrod_1806
Chlamydophila caviae	CCA_00059	CCA_00060
Clostridium cellulovorans	Clocel_2940	Clocel_2941
Clostridium cellulolyticum	Ccel_2118	Ccel_2117
Clostridium cellulosi	CCDG5_1881	CCDG5_1880
Clostridium carboxidivorans	Ccar_00530	Ccar_00535
Clostridium carboxidivorans	Ccar_04255	Ccar_04260
Clostridium clariflavum	Clocl_3366	Clocl_3367
Cryptobacterium curtum	Ccur_04040	Ccur_04030
Corallococcus coralloides	COCOR_04743	COCOR_04744
Caldisericum exile	CSE_12730	CSE_12740
Chlamydia felis Fe/C-56	CF0945	CF0944
Cellulomonas flavigena	Cfla_2818	Cfla_2817
Collimonas fungivorans	CFU_2481	CFU_2480
Cellulomonas gilvus	Celgi_2642	Celgi_2643
Cyanobium gracile	Cyagr_2677	Cyagr_2678
Cupriavidus gilardii	CR3_1087	CR3_1088
Chryseobacterium gallinarum	OK18_13080	OK18_13085
Chlamydia gallinacea	M787_002225	M787_002230
Caldicellulosiruptor hydrothermalis	Calhy_1252	Calhy_1251
Chlamydia psittaci 6BC	CPSIT_0071	CPSIT_0072
Carboxydothermus hydrogenoformans	CHY_0597	CHY_0596
Chryseobacterium sp. StRB126	CHSO_1320	CHSO_1321
Chryseobacterium sp. IHB B 17019	ATE47_02555	ATE47_02560
Cellvibrio japonicus	CJA_0482	CJA_0481
Caldicellulosiruptor kristjanssonii	Calkr_1469	Calkr_1470
Clostridium kluyveri DSM 555	CKL_1353	CKL_1354
Caldicellulosiruptor kronotskyensis	Calkro_1226	Calkro_1225
Clostridium ljungdahlii	CLJU_c12440	CLJU_c12450
Mageeibacillus indolicus	HMPREF0868_0831	HMPREF0868_0830
Chlamydophila pneumoniae LPCoLN	CPK_ORF00082	CPK_ORF00081
Caldivirga maquilingensis	Cmaq_1965	Cmaq_1966

Chlamydia muridarum Nigg	TC_0063	TC_0064
Cupriavidus necator N-1	CNE_1c21790	CNE_1c21780
Calditerrivibrio nitroreducens	Calni_1127	Calni_1128
Caldicellulosiruptor obsidiansis	COB47_1070	COB47_1069
Colwellia sp. MT41	CMT41_12550	CMT41_12545
Coprococcus sp. ART55/1	CCU_16630	CCU_16640
Caldicellulosiruptor owensensis	Calow_1263	Calow_1264
Clostridium pasteurianum BC1	Clopa_3800	Clopa_3801
Clostridium pasteurianum BC1	Clopa_4475	Clopa_4474
Clostridium perfringens 13	CPE0582	CPE0583
Chitinophaga pinensis	Cpin_0179	Cpin_0180
Chlamydia pecorum E58	G5S_0366	G5S_0367
Chlamydia pneumoniae CWL029	CPn0681	CPn0680
Collimonas pratensis	CPter91_3226	CPter91_3225
Colwellia psychrerythraea	CPS_3638	CPS_3637
Chromohalobacter salexigens	Csal_0711	Csal_0712
Clostridium saccharobutylicum	CLSA_c23780	CLSA_c23770
Caldicellulosiruptor saccharolyticus	Csac_2031	Csac_2032
Caulobacter segnis	Cseg_2924	Cseg_2925
Clostridium scatologenes	CSCA_2064	CSCA_2063
Clostridium scatologenes	CSCA_2853	CSCA_2852
Clostridium saccharoperbutylacetonicum	Cspa_c35370	Cspa_c35360
Clostridium stercorarium subsp. stercorarium DSM 8532	Cst_c11110	Cst_c11100
Ruminiclostridium thermocellum ATCC 27405	Cthe_3001	Cthe_3000
Cupriavidus taiwanensis	RALTA_A1818	RALTA_A1817
Chlamydia trachomatis D/UW-3/CX	CT_691	CT_692
Comamonas testosteroni CNB-2	CtCNB1_0893	CtCNB1_0894
Clostridium tyrobutyricum	CTK_C18550	CTK_C18540
Conexibacter woesei	Cwoe_2888	Cwoe_2887
Cycloclasticus sp. P1	Q91_0313	Q91_0314
Deinococcus actinosclerus	AUC44_02875	AUC44_02870
Delftia acidovorans	Daci_5349	Daci_5348
Desulfotomaculum acetoxidans	Dtox_2384	Dtox_2385
Desulfovibrio africanus	Desaf_1055	Desaf_1056
Desulfosporosinus acidiphilus	Desaci_4323	Desaci_4322
Desulfatibacillum alkenivorans	Dalk_4121	Dalk_4122
Dechloromonas aromatica	Daro_0678	Daro_0679
Dechloromonas aromatica	Daro_1122	Daro_1123
Desulfovibrio aespoeensis	Daes_0822	Daes_0821
Desulfotomaculum nigrificans	Desca_2146	Desca_2147
Deinococcus soli	SY84_11620	SY84_11615
Desulfovibrio alaskensis	Dde_3780	Dde_3781
Desulfitobacterium dehalogenans	Desde_3733	Desde_3732

Desulfitobacterium dichloroeliminans	Desdi_3072	Desdi_3071
Deinococcus deserti	Deide_03420	Deide_03410
Desulfovibrio desulfuricans ATCC 27774	Ddes_1961	Ddes_1960
Dehalobacter sp. CF	DCF50_p1845	DCF50_p1844
Dehalobacter sp. DCA	DHBDCA_p1834	DHBDCA_p1833
Desulfovibrio sp. J2	AWY79_05305	AWY79_05300
Dehalococcoides mccartyi 195	DET0033	DET0034
Dehalogenimonas sp. WBC-2	DGWBC_0234	DGWBC_0235
Desulfovibrio fairfieldensis	AXF13_11015	AXF13_11010
Deinococcus geothermalis	Dgeo_1747	Dgeo_1748
Desulfovibrio gigas	DGI_2144	DGI_2143
Deinococcus gobiensis	DGo_CA0518	DGo_CA0517
Desulfovibrio hydrothermalis	DESAM_20632	DESAM_20631
Acidovorax ebreus	Dtpsy_2707	Dtpsy_2706
Desulfurispirillum indicum	Selin_1028	Selin_1029
Dokdonella koreensis	I596_3010	I596_3009
Desulfotomaculum kuznetsovii	Desku_0374	Desku_0375
Dehalogenimonas lykanthroporepellens	Dehly_0087	Dehly_0086
Desulfovibrio magneticus	DMR_08640	DMR_08650
Desulfosporosinus meridiei	Desmer_4203	Desmer_4202
Deinococcus maricopensis	Deima_2279	Deima_2280
Desulfitobacterium metallireducens	DESME_13545	DESME_13540
Candidatus Desulfococcus oleovorans	Dole_3224	Dole_3225
Desulfosporosinus orientis	Desor_3060	Desor_3059
Desulfosporosinus orientis	Desor_5227	Desor_5226
Desulfovibrio piezophilus	BN4_11640	BN4_11641
Dialister pneumosintes	BCB69_01245	BCB69_01250
Desulfobulbus propionicus	Despr_1330	Despr_1331
Desulfotalea psychrophila	DP2627	DP2628
Deinococcus proteolyticus	Deipr_1226	Deipr_1227
Deinococcus radiodurans	DR_0924	DR_0925
Desulfotomaculum reducens	Dred_2657	Dred_2658
Dehalobacter restrictus	DEHRE_02825	DEHRE_02830
Desulfotomaculum ruminis	Desru_0952	Desru_0951
Desulfovibrio salexigens	Desal_0093	Desal_0094
Desulfocapsa sulfexigens	UWK_00038	UWK_00037
Dechlorosoma suillum	Dsui_1844	Dsui_1845
Dechlorosoma suillum	Dsui_3024	Dsui_3023
Deinococcus swuensis	QR90_13120	QR90_13115
Desulfitobacterium hafniense Y51	DSY3186	DSY3185
Desulfomonile tiedjei	Desti_4215	Desti_4216
Desulfomonile tiedjei	Desti_4364	Desti_4363
Desulfobacula toluolica	TOL2_C25580	TOL2_C25590

Dehalococcoides sp. UCH007	UCH007_00310	UCH007_00320
Dermabacter vaginalis	DAD186_02710	DAD186_02700
Desulfovibrio vulgaris Hildenborough	DVU2305	DVU2306
Peptoclostridium acidaminophilum	EAL2_c15530	EAL2_c15520
Ensifer adhaerens OV14	OV14_1013	OV14_1014
Exiguobacterium antarcticum	Eab7_2712	Eab7_2713
Elizabethkingia anophelis NUHP1	BD94_1433	BD94_1432
Enterococcus casseliflavus	ECBG_02215	ECBG_02214
Edwardsiella sp. LADL05-105	AAZ33_02160	AAZ33_02155
Edwardsiella sp. EA181011	QY76_00585	QY76_00590
Enterococcus faecalis V583	EF2441	EF2442
Enterococcus gallinarum	AL523_12350	AL523_12345
Ethanoligenens harbinense	Ethha_0025	Ethha_0026
Edwardsiella ictaluri	NT01EI_0534	NT01EI_0533
Elizabethkingia miricola	VO54_01105	VO54_01106
Eggerthella lenta	Elen_2716	Elen_2717
Elusimicrobium minutum	Emin_0044	Emin_0043
Exiguobacterium sibiricum	Exig_2900	Exig_2901
Enterococcus silesiacus	ATZ33_05500	ATZ33_05505
Edwardsiella piscicida C07-087	ETAC_02260	ETAC_02255
Edwardsiella anguillarum	ETEE_2225	ETEE_2224
Edwardsiella tarda EIB202	ETAE_0464	ETAE_0463
Exiguobacterium sp. MH3	U719_15815	U719_15820
Eggerthella sp. YY7918	EGYY_25890	EGYY_25900
Fictibacillus arsenicus	ABE41_007715	ABE41_007710
Flavobacteriaceae bacterium 3519-10	FIC_00734	FIC_00735
Ferrimonas balearica	Fbal_0553	Fbal_0552
Flavobacterium branchiophilum	FBFL15_1359	FBFL15_1360
Francisella cf. novicida Fx1	FNFX1_0081	FNFX1_0082
Flavobacterium columnare	FCOL_06085	FCOL_06065
Fimbriimonas ginsengisoli	OP10G_1337	OP10G_1338
Allofrancisella guangzhouensis	SD28_00940	SD28_00935
Flavobacterium indicum	KQS_05550	KQS_05555
Flavobacterium johnsoniae	Fjoh_2577	Fjoh_2578
Flavisolibacter sp. LCS9	SY85_18970	SY85_18965
Fervidobacterium nodosum	Fnod_0493	Fnod_0492
Fervidobacterium pennivorans	Ferpe_0820	Ferpe_0821
Fictibacillus phosphorivorans	ABE65_006670	ABE65_006665
Flavobacterium psychrophilum JIP02/86	FP1001	FP1000
Francisella sp. FSC1006	LO80_06615	LO80_06620
Fluviicola taffensis	Fluta_2325	Fluta_2324
Geoglobus acetivorans	GACE_1915	GACE_1914
Glaciecola sp. 4H-3-7+YE-5	Glaag_1031	Glaag_1032

Geoglobus ahangari	GAH_00740	GAH_00741
Gallibacterium anatis	UMN179_01864	UMN179_01863
Geobacter anodireducens	A2G06_01730	A2G06_01735
Gemmatimonas aurantiaca	GAU_2374	GAU_2373
Gemmatirosa kalamazoonesis	J421_3846	J421_3845
Geobacter bemidjiensis	Gbem_3782	Gbem_3781
Geobacter daltonii FRC-32	Geob_0604	Geob_0605
Gemmata sp. SH-PL17	VT84_11710	VT84_11715
Grimontia hollisae	AL542_11945	AL542_11940
Geobacillus kaustophilus	GK2946	GK2945
Geobacter lovleyi	Glov_0586	Glov_0587
Geobacter metallireducens	Gmet_3132	Gmet_3131
Gamma proteobacterium HdN1	HDN1F_12670	HDN1F_12680
Geobacter pickeringii	GPICK_14340	GPICK_14335
Paraglaciecola psychrophila	C427_1521	C427_1522
Geobacillus stearothermophilus	GT50_06365	GT50_06360
Geobacillus subterraneus	GS3922_01460	GS3922_01465
Geobacter sulfurreducens PCA	GSU0388	GSU0389
Geobacillus thermoglucosidasius	Geoth_0618	Geoth_0619
Geobacillus thermodenitrificans	GTNG_2898	GTNG_2897
Geobacter uraniireducens	Gura_4136	Gura_4135
Geobacillus sp. WCH70	GWCH70_2877	GWCH70_2876
Halotalea alkalilenta	A5892_06915	A5892_06920
Halomonas sp. KO116	KO116_03549	KO116_03548
Halomonas sp. R57-5	HALO3805	HALO3804
Haemophilus parasuis SH0165	HAPS_0690	HAPS_0689
Herminiimonas arsenicoxydans	HEAR2157	HEAR2156
Hafnia alvei	AT03_03960	AT03_03955
Halomonas chromatireducens	LOKO_02438	LOKO_02437
Halomonas campaniensis	FF32_09900	FF32_09905
Hyphomicrobium denitrificans ATCC 51888	Hden_0820	Hden_0819
Haemophilus ducreyi	HD_1371	HD_1370
Halomonas elongata	HELO_1615	HELO_1616
Halorhodospira halophila	Hhal_1052	Hhal_1053
Herbaspirillum hiltneri	F506_10440	F506_10445
Halomonas huangheensis	AR456_03395	AR456_03400
Haemophilus influenzae Rd KW20 (serotype d)	HI1603	HI1604
Heliobacterium modesticaldum	HM1_2747	HM1_2746
Hippea maritima	Hipma_0898	Hipma_0897
Haemophilus parainfluenzae	PARA_14140	PARA_14130
Herbaspirillum rubrisubalbicans	Hrubri_1961	Hrubri_1962
Herbinix luporum	SD1D_1863	SD1D_1862
Herbaspirillum seropedicae SmR1	Hsero_2080	Hsero_2081

Haemophilus somnus 129PT	HS_1190	HS_1191
Ignavibacterium album	IALB_2613	IALB_2614
Intrasporangium calvum	Intca_2958	Intca_2957
Isoptericola dokdonensis	I598_1986	I598_1985
Idiomarina loihiensis L2TR	IL2208	IL2207
Isoptericola variabilis	Isova_0375	Isova_0376
Janthinobacterium agaricidamnosum	GJA_3594	GJA_3593
Kribbella flavida	Kfla_0949	Kfla_0950
Kribbella flavida	Kfla_3847	Kfla_3848
Kocuria flava	AS188_06155	AS188_06150
Kangiella geojedonensis	TQ33_0421	TQ33_0420
Kangiella koreensis	Kkor_2151	Kkor_2152
Kosmotoga olearia	Kole_0539	Kole_0540
Kosmotoga olearia	Kole_0887	Kole_0888
Kosmotoga pacifica	IX53_05230	IX53_05225
Kineococcus radiotolerans	Krad_0870	Krad_0871
Kangiella sediminilitoris	KS2013_439	KS2013_438
Kitasatospora setae	KSE_43450	KSE_43440
Kurthia sp. 11kri321	ASO14_232	ASO14_231
Kurthia sp. 11kri321	ASO14_2331	ASO14_2330
Lysobacter antibioticus	LA76x_1817	LA76x_1818
Lachnoclostridium sp. YL32	A4V08_22730	A4V08_22725
Lacimicrobium alkaliphilum	AT746_14720	AT746_14715
Leptothrix cholodnii	Lcho_0677	Lcho_0678
Lysobacter capsici	LC55x_1737	LC55x_1738
Lysobacter enzymogenes	GLE_3501	GLE_3500
Legionella fallonii	LFA_3402	LFA_3403
Lysinibacillus fusiformis	HR49_02310	HR49_02305
Lysinibacillus fusiformis	HR49_14495	HR49_14490
Lysobacter gummosus	LG3211_3438	LG3211_3437
Lysinibacillus varians	T479_09575	T479_09580
Lysinibacillus varians	T479_21805	T479_21810
Legionella hackeliae	LHA_2688	LHA_2689
Limnohabitans sp. 103DPR2	L103DPR2_02379	L103DPR2_02378
Listeria innocua	lin2350	lin2351
Listeria ivanovii subsp. ivanovii PAM 55	LIV_2232	LIV_2233
Legionella longbeachae	LLO_2741	LLO_2742
Labilithrix luteola	AKJ09_05269	AKJ09_05270
Listeria monocytogenes EGD-e	lmo2248	lmo2249
Luteipulveratus mongoliensis	VV02_16730	VV02_16735
Legionella oakridgensis	Loa_02651	Loa_02652
Legionella pneumophila subsp. pneumophila Philadelphia 1	lpg0584	lpg0583

Listeria seeligeri	lse_2228	lse_2229
Lysinibacillus sphaericus	Bsph_2363	Bsph_2364
Lysinibacillus sphaericus	Bsph_4376	Bsph_4377
Listeria welshimeri	lwe2264	lwe2265
Methanosarcina acetivorans	MA_2934	MA_2935
Methanococcus aeolicus	Maeo_0152	Maeo_0151
Magnetospirillum magneticum	amb4471	amb4472
Magnetospira sp. QH-2	MGMAQ_2282	MGMAQ_2281
Magnetospirillum sp. XM-1	XM1_0765	XM1_0766
Marinobacterium sp. ST58-10	A8C75_01205	A8C75_01200
Massilia sp. WG5	AM586_21085	AM586_21080
Micromonospora aurantiaca	Micau_5731	Micau_5730
Methanoculleus bourgensis	BN140_2330	BN140_2331
Methanoregula boonei	Mboo_0763	Mboo_0762
Methanoregula boonei	Mboo_0763	Mboo_0764
Methanoregula boonei	Mboo_0765	Mboo_0764
Methanococcoides burtonii	Mbur_2044	Mbur_2045
Macrococcus caseolyticus	MCCL_0402	MCCL_0403
Methylomonas sp. DH-1	AYM39_17445	AYM39_17440
Methylomonas denitrificans	JT25_003660	JT25_003665
Methylomonas denitrificans	JT25_013525	JT25_013530
Megasphaera elsdenii	MELS_0578	MELS_0577
Megasphaera elsdenii	MELS_0880	MELS_0881
Methanosarcina sp. WH1	MSWH1_1343	MSWH1_1344
Methylotenera versatilis	M301_1984	M301_1985
Methanosarcina sp. Kolksee	MSKOL_2454	MSKOL_2453
Methanoculleus marisnigri	Memar_1039	Memar_1038
Methanosarcina sp. WWM596	MSWHS_1429	MSWHS_1430
Methylobacterium sp. 4-46	M446_5971	M446_5972
Methanosarcina sp. MTP4	MSMTP_1042	MSMTP_1043
Methanosarcina sp. MTP4	MSMTP_1739	MSMTP_1738
Methylophilus sp. TWE2	ACJ67_05745	ACJ67_05740
Methanocaldococcus fervens	Mefer_1425	Mefer_1426
Methanoregula formicica	Metfor_1540	Metfor_1541
Methanoregula formicica	Metfor_2073	Metfor_2074
Methanoregula formicica	Metfor_2075	Metfor_2074
Methanoregula formicica	Metfor_2075	Metfor_2076
Methanocaldococcus sp. FS406-22	MFS40622_1041	MFS40622_1042
Myxococcus fulvus	LILAB_19115	LILAB_19110
Magnetococcus marinus	Mmc1_0228	Mmc1_0229
Magnetospirillum gryphiswaldense	MGMSRv2__2522	MGMSRv2__2523
Megamonas hypermegale	MHY_10360	MHY_10370

Megamonas hypermegale	MHY_28810	MHY_28820
Methanosarcina horonobensis	MSHOH_1782	MSHOH_1781
Mannheimia haemolytica USDA-ARS-USMARC-185	D648_11730	D648_11720
Methanospirillum hungatei	Mhun_1022	Mhun_1021
Methanospirillum hungatei	Mhun_1022	Mhun_1023
Methanospirillum hungatei	Mhun_1024	Mhun_1023
Methanomethylovorans hollandica	Metho_2270	Metho_2271
Candidatus Beckwithbacteria bacterium GW2011_GWC1_49_16	UY43_C0001G0238	UY43_C0001G0237
Methanotorris igneus	Metig_1180	Metig_1181
Micromonospora sp. L5	ML5_2765	ML5_2766
Mitsuaria sp. 7	ABE85_02875	ABE85_02880
Methanocaldococcus jannaschii	MJ_0629	MJ_0630
Methanocaldococcus bathoardescens	JH146_0850	JH146_0851
Methanopyrus kandleri	MK0562	MK0561
Methanocorpusculum labreanum	Mlab_0997	Mlab_0998
Mesorhizobium loti	mlr8369	mlr8370
Mesorhizobium loti	msr8368	mlr8370
Methanosarcina lacustris	MSLAZ_1645	MSLAZ_1644
Methanosarcina mazei Go1	MM_3231	MM_3230
Modestobacter marinus	MODMU_1282	MODMU_1281
Methylotenera mobilis	Mmol_1688	Mmol_1689
Methanococcoides methylutens	MCMEM_0260	MCMEM_0261
Janthinobacterium sp. Marseille	mma_1306	mma_1307
Methylobacterium nodulans	Mnod_8017	Mnod_8018
Massilia sp. NR 4-1	ACZ75_26115	ACZ75_26120
Myroides odoratimimus	AS202_15865	AS202_15870
Methanothermococcus okinawensis	Metok_1127	Metok_1128
Methylobacterium oryzae	MOC_0987	MOC_0986
Methanocella paludicola	MCP_0107	MCP_0108
Mesotoga prima	Theba_2711	Theba_2712
Methanolacinia petrolearia	Mpet_0495	Mpet_0494
Methanolacinia petrolearia	Mpet_0830	Mpet_0831
Methanolacinia petrolearia	Mpet_1965	Mpet_1964
Candidatus Methanosphaerula palustris	Mpal_1967	Mpal_1966
Methylibium petroleiphilum	Mpe_A3016	Mpe_A3015
Marichromatium purpuratum	MARPU_02060	MARPU_02055
Myroides profundi	MPR_2640	MPR_2639
Methanolobus psychrophilus	Mpsy_2012	Mpsy_2011
Marinitoga piezophila	Marpi_0064	Marpi_0063
Marinitoga piezophila	Marpi_0648	Marpi_0647
Methylobacterium radiotolerans	Mrad2831_0776	Mrad2831_0775
Methylocystis sp. SC2	BN69_0197	BN69_0198

Myxococcus stipitatus	MYSTI_02630	MYSTI_02629
Methanosarcina siciliae C2J	MSSAC_1552	MSSAC_1551
Mannheimia succiniciproducens	MS2309	MS2310
Meiothermus silvanus	Mesil_1084	Mesil_1083
Moorella thermoacetica	Moth_1236	Moth_1237
Microbulbifer thermotolerans	A3224_13965	A3224_13970
Methanosarcina thermophila CHTI-55	MSTHC_0479	MSTHC_0478
Mucilaginibacter sp. PAMC 26640	A0256_04055	A0256_04050
Methanosarcina vacuolata	MSVAZ_2441	MSVAZ_2440
Mannheimia varigena USDA-ARS-USMARC-1296	X808_9770	X808_9760
Mannheimia sp. USDA-ARS-USMARC-1261	X781_13040	X781_13050
Moritella viscosa	MVIS_4090	MVIS_4091
Methanocaldococcus vulcanius	Metvu_1552	Metvu_1553
Myxococcus xanthus	MXAN_2281	MXAN_2280
Myxococcus hansupus	A176_004769	A176_004770
Myroides sp. A21	MYRA21_0008	MYRA21_0007
Novosphingobium aromaticivorans	Saro_0842	Saro_0843
Candidatus Nitrosopelagicus brevis	T478_0952	T478_0951
Nocardioides sp. JS614	Noca_4258	Noca_4259
Nitrospira defluvii	NIDE0855	NIDE0856
Nocardioides dokdonensis	I601_1711	I601_1710
Candidatus Nitrososphaera evergladensis	NTE_00866	NTE_00865
Candidatus Nitrososphaera gargensis	Ngar_c24690	Ngar_c24680
Neorhizobium galegae bv. officinalis bv. officinalis HAMBI 1141	RG1141_CH19140	RG1141_CH19150
Nitrobacter hamburgensis	Nham_1677	Nham_1678
Niabella sp. BS26	A8C56_21695	A8C56_21690
Candidatus Nitrosopumilus sp. AR2	NSED_02415	NSED_02420
Niastella koreensis	Niako_6887	Niako_6886
Candidatus Nitrosopumilus koreensis	NKOR_02485	NKOR_02490
Nakamurella multipartita	Namu_3507	Namu_3506
Nitrospira moscoviensis	NITMOv2_0712	NITMOv2_0713
Novosphingobium pentaromativorans	JI59_13390	JI59_13395
Novosphingobium sp. PP1Y	PP1Y_AT33953	PP1Y_AT33945
Niabella soli	NIASO_15190	NIASO_15195
Nitrososphaera viennensis	NVIE_026490	NVIE_026480
Nitrobacter winogradskyi	Nwi_2032	Nwi_2031
Opitutaceae bacterium TAV5	OPIT5_05540	OPIT5_05545
Oligotropha carboxidovorans OM5 (Mississippi)	OCAR_6539	OCAR_6540
Oceanimonas sp. GK1	GU3_15655	GU3_15660
Olsenella sp. oral taxon 807	ADJ70_03850	ADJ70_03855
Olsenella uli	Olsu_1174	Olsu_1173
Olsenella uli	Olsu_1390	Olsu_1391

Oceanithermus profundus	Ocepr_0031	Ocepr_0032
Opitutus terrae	Oter_2500	Oter_2499
Opitutus terrae	Oter_2553	Oter_2552
Ottowia sp. oral taxon 894	ADJ79_10545	ADJ79_10540
Pyrococcus abyssi	PAB0929	PAB0927
Pseudomonas aeruginosa PAO1	PA5208	PA5207
Paenibacillus sp. FSL P4-0081	P40081_03865	P40081_03870
Pyrobaculum aerophilum	PAE3187	PAE3188
Pandoraea apista	SG18_18605	SG18_18610
Pseudoalteromonas atlantica	Patl_3320	Patl_3319
Oblitimonas alkaliphila	AKN87_09380	AKN87_09375
Paenibacillus borealis	PBOR_03570	PBOR_03575
Polyangium brachysporum	AAW51_3758	AAW51_3757
Paenibacillus beijingensis	VN24_06750	VN24_06745
Pseudomonas balearica	CL52_19505	CL52_19500
Paenibacillus bovis	AR543_06215	AR543_06220
Pedobacter cryoconitis	AY601_3925	AY601_3926
Candidatus Protochlamydia amoebophila	pc0098	pc0099
Paenibacillus durus	PDUR_03735	PDUR_03740
Pseudonocardia dioxanivorans	Psed_2917	Psed_2916
Pedobacter sp. PACM 27299	AQ505_25120	AQ505_25125
Pandoraea faecigallinarum	AB870_21210	AB870_21215
Pelosinus fermentans	JBW_02840	JBW_02839
Pyrococcus furiosus DSM 3638	PF1021	PF1020
Pseudomonas fulva	Psefu_0508	Psefu_0509
Photobacterium gaetbulicola	H744_2c3331	H744_2c3332
Paenibacillus graminis	PGRAT_03875	PGRAT_03880
Pseudoalteromonas haloplanktis	PSHAa0313	PSHAa0314
Pedobacter heparinus	Phep_0191	Phep_0190
Pyrococcus horikoshii	PH0637	PH0640
Pseudothermotoga hypogea	AJ81_01075	AJ81_01070
Paenibacillus sp. IHBB 10380	UB51_00215	UB51_00220
Psychromonas ingrahamii	Ping_0181	Ping_0180
Paenibacillus sp. JDR-2	Pjdr2_0875	Pjdr2_0876
Paucibacter sp. KCTC 42545	AT984_10215	AT984_10210
Planococcus kocurii	AUO94_12095	AUO94_12090
Planococcus sp. L10.15	I858_002940	I858_002935
Paenibacillus larvae	ERIC2_c14060	ERIC2_c14050
Petrotoga mobilis	Pmob_0285	Pmob_0286
Paenibacillus mucilaginosus KNP414	KNP414_02829	KNP414_02828
Pasteurella multocida subsp. multocida Pm70	PM0250	PM0249
Pseudomonas mendocina ymp	Pmen_0351	Pmen_0352
Polaromonas naphthalenivorans	Pnap_1151	Pnap_1152

Polynucleobacter necessarius subsp. necessarius	Pnec_0558	Pnec_0559
Protochlamydia naegleriophila	PNK_2305	PNK_2304
Paenibacillus naphthalenovorans	IJ22_22570	IJ22_22580
Pandoraea norimbergensis	AT302_22675	AT302_22670
Polynucleobacter asymbioticus	Pnuc_1405	Pnuc_1404
Paenibacillus odorifer	PODO_03425	PODO_03430
Polaromonas sp. JS666	Bpro_1686	Bpro_1687
Paenibacillus sp. 32O-W	IJ21_46280	IJ21_46270
Pandoraea oxalativorans	MB84_15125	MB84_15130
Palaeococcus pacificus	PAP_10240	PAP_10245
Pelobacter propionicus	Ppro_3584	Ppro_3585
Paenibacillus peoriae	ABE82_04155	ABE82_04160
Pandoraea pnomenusa 3kgm	U875_13355	U875_13350
Paludibacter propionicigenes	Palpr_2883	Palpr_2882
Photobacterium profundum	PBPRA0444	PBPRA0443
Pseudomonas pseudoalcaligenes	BN5_0271	BN5_0272
Pandoraea pulmonicola	RO07_19515	RO07_19520
Paenibacillus polymyxa E681	PPE_00804	PPE_00805
Paenibacillus riograndensis	PRIO_0804	PRIO_0805
Pseudoalteromonas rubra	AT705_14660	AT705_14665
Pseudomonas stutzeri A1501	PST_0272	PST_0273
Paenibacillus sabinae	PSAB_03230	PSAB_03235
Pimelobacter simplex	KR76_22940	KR76_22945
Pseudoalteromonas sp. SM9913	PSM_A2780	PSM_A2779
Pseudopedobacter saltans	Pedsa_2986	Pedsa_2985
Pandoraea sputorum	NA29_17705	NA29_17700
Paenibacillus stellifer	PSTEL_04040	PSTEL_04045
Pedobacter steynii	BFS30_10915	BFS30_10910
Pseudoxanthomonas suwonensis 11-1	Psesu_1862	Psesu_1861
Pseudarthrobacter sulfonivorans	AU252_08605	AU252_08610
Psychromonas sp. CNPT3	PCNPT3_12585	PCNPT3_12590
Paenibacillus terrae	HPL003_11955	HPL003_11960
Pelotomaculum thermopropionicum	PTH_2036	PTH_2037
Parageobacillus thermoglucosidans	AOT13_00125	AOT13_00120
Pseudoalteromonas translucida	PTRA_a0333	PTRA_a0334
Picrophilus torridus	PTO1328	PTO1327
Pelosinus sp. UFO1	UFO1_1494	UFO1_1495
Pelosinus sp. UFO1	UFO1_3189	UFO1_3190
Parachlamydia acanthamoebae	PUV_02710	PUV_02720
Parachlamydia acanthamoebae	PUV_06440	PUV_06450
Pandoraea vervacti	UC34_10760	UC34_10765
Pyrococcus yayanosii	PYCH_10440	PYCH_10430
Pyrococcus sp. NCB100	TQ32_04665	TQ32_04670

Pyrococcus sp. NA2	PNA2_1258	PNA2_1259
Pyrococcus sp. ST04	Py04_1115	Py04_1114
Phenylobacterium zucineum	PHZ_c3313	PHZ_c3314
Riemerella anatipestifer ATCC 11845 = DSM 15868	Riean_0508	Riean_0507
Rubrivivax benzoatilyticus	RBXJA2T_05418	RBXJA2T_05423
Methanocella arvoryzae	RCIX291	RCIX290
Roseateles depolymerans	RD2015_3627	RD2015_3626
Rhizobium etli CFN 42	RHE_CH00947	RHE_CH00946
Ralstonia eutropha JMP134	Reut_A2001	Reut_A2000
Rhodoferax ferrireducens	Rfer_1398	Rfer_1399
Rhizobium gallicum	RGR602_PC00090	RGR602_PC00089
Rubrivivax gelatinosus	RGE_13430	RGE_13440
Rhodococcus sp. B7740	NY08_4465	NY08_4466
Sinorhizobium fredii NGR234	NGR_c35010	NGR_c35020
Rhodoplanes sp. Z2-YC6860	RHPLAN_08640	RHPLAN_08630
Ralstonia insidiosa	ACS15_2057	ACS15_2056
Rhizobium sp. IRBG74	BN877_II1720	BN877_II1719
Rhizobium leguminosarum bv. viciae 3841	pRL100342	pRL100343
Candidatus Ruthia magnifica (Calyptogena magnifica)	Rmag_0124	Rmag_0125
Cupriavidus metallidurans	Rmet_1974	Rmet_1973
Ralstonia mannitolilytica	TK49_10710	TK49_10715
Rothia mucilaginosa	RMDY18_01780	RMDY18_01770
Rhodopseudomonas palustris CGA009	RPA2280	RPA2281
Ralstonia pickettii 12J	Rpic_1187	Rpic_1188
Renibacterium salmoninarum	RSal33209_3171	RSal33209_3172
Ralstonia solanacearum CFBP2957	RCFBP_20114	RCFBP_20113
Ralstonia solanacearum GMI1000	RSc1312	RSc1313
Rummeliibacillus stabekisii	ATY39_01690	ATY39_01695
Rummeliibacillus stabekisii	ATY39_12950	ATY39_12955
Ramlibacter tataouinensis	Rta_12050	Rta_12060
Staphylococcus aureus RF122	SAB0612	SAB0613
Stenotrophomonas acidaminiphila	AOT14_14500	AOT14_14510
Simiduia agarivorans	M5M_13625	M5M_13620
Staphylococcus agnetis	EP23_11750	EP23_11755
Sulfolobus acidocaldarius DSM 639	Saci_0266	Saci_0265
Sphingopyxis alaskensis	Sala_0716	Sala_0717
Streptomyces albus J1074	XNR_3243	XNR_3244
Streptomyces albulus NK660	DC74_4662	DC74_4663
Sandaracinus amylolyticus	DB32_006071	DB32_006070
Salinispora arenicola	Sare_0325	Sare_0326
Syntrophus aciditrophicus	SYN_00205	SYN_00206
Syntrophus aciditrophicus	SYN_00205	SYN_03078

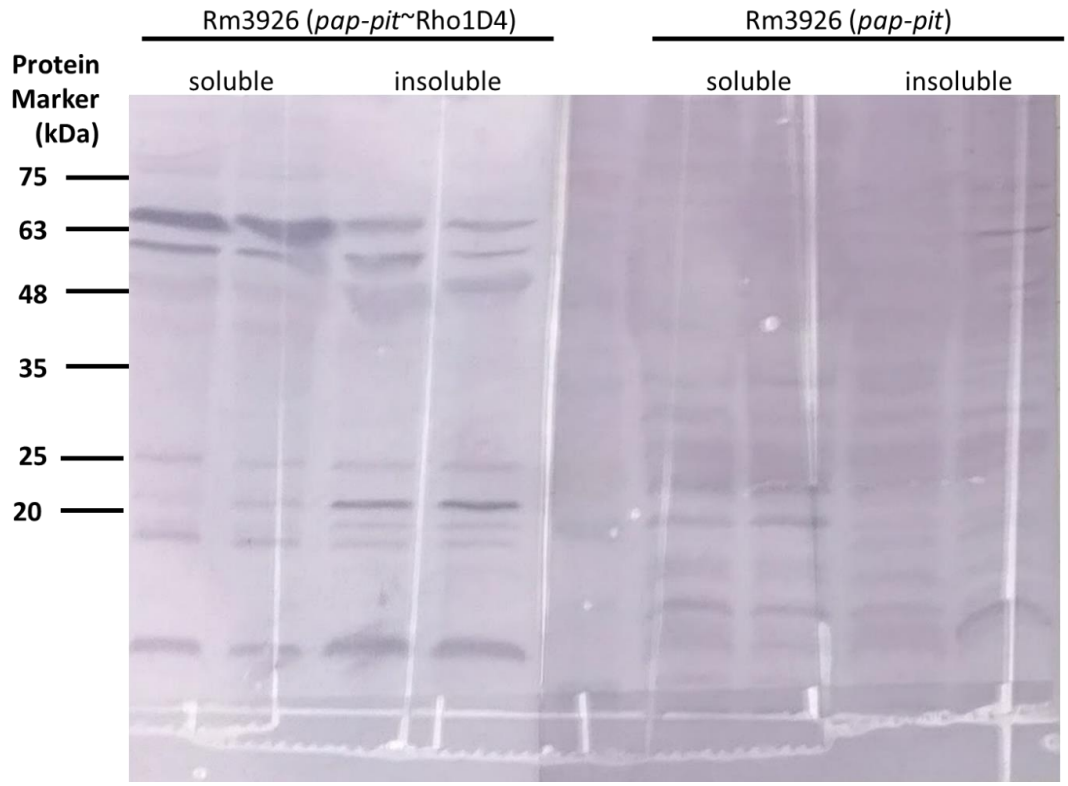
Syntrophus aciditrophicus	SYN_00386	SYN_00387
Sinomonas atrocyanea	SA2016_0300	SA2016_0299
Staphylococcus aureus subsp. aureus N315 (MRSA/VSSA)	SA0618	SA0619
Sulfobacillus acidophilus TPY	TPY_1552	TPY_1553
Shewanella amazonensis	Sama_2673	Sama_2674
Sphingobium baderi	ATN00_19815	ATN00_19820
Streptomyces bingchenggensis	SBI_05100	SBI_05099
Shewanella baltica OS155	Sbal_3455	Sbal_3456
Staphylococcus carnosus	SCA_0306	SCA_0307
Staphylococcus capitis	AYP1020_0029	AYP1020_0030
Streptomyces scabiei	SCAB_49361	SCAB_49351
Sphaerochaeta coccoides	Spico_0163	Spico_0164
Sphingobium chlorophenolicum	Sphch_2570	Sphch_2569
Streptomyces collinus	B446_18100	B446_18105
Sorangium cellulosum So ce 56	sce3331	sce3330
Sorangium cellulosum So ce 56	sce3424	sce3425
Solitalea canadensis	Solca_2748	Solca_2747
Streptomyces coelicolor	SCO4137	SCO4138
Streptomyces cattleya NRRL 8057 = DSM 46488	SCAT_2577	SCAT_2576
Shewanella denitrificans	Sden_0692	Sden_0691
Sulfuricella denitrificans	SCD_n00184	SCD_n00183
Streptomyces davawensis	BN159_4764	BN159_4765
Selenomonas sp. oral taxon 478	ADJ74_09815	ADJ74_09820
Staphylococcus epidermidis ATCC 12228	SE0431	SE0432
Staphylococcus equorum	SE1039_04880	SE1039_04890
Serinicoccus sp. JLT9	SGUI_0475	SGUI_0474
Streptomyces pratensis	Sfla_2908	Sfla_2907
Streptomyces fulvissimus	SFUL_3914	SFUL_3915
Shewanella frigidimarina	Sfri_0565	Sfri_0566
Shewanella frigidimarina	Sfri_3884	Sfri_3885
Syntrophobacter fumaroxidans	Sfum_2274	Sfum_2275
Streptomyces globisporus	WQO_15965	WQO_15960
Sphingopyxis granuli	SGRAN_1407	SGRAN_1408
Streptomyces griseus	SGR_3928	SGR_3929
Streptomyces glaucescens	SGLAU_16005	SGLAU_16010
Syntrophobotulus glycolicus	Sgly_1671	Sgly_1670
Staphylococcus haemolyticus JCSC1435	SH2231	SH2230
Sulfuritalea hydrogenivorans	SUTH_02217	SUTH_02216
Shewanella sp. MR-4	Shewmr4_3119	Shewmr4_3120
Sphingobacterium sp. 21	Sph21_1370	Sph21_1371
Shewanella halifaxensis	Shal_3582	Shal_3583
Sphingobacterium sp. ML3W	KO02_23380	KO02_23385
Staphylococcus hyicus	SHYC_10415	SHYC_10410

Streptomyces hygroscopicus subsp. jinggangensis 5008	SHJG_4940	SHJG_4941
Shinella sp. HZN7	shn_07725	shn_07720
Sulfolobus islandicus L.S.2.15	LS215_0279	LS215_0280
Solibacillus silvestris StLB046	SSIL_0366	SSIL_0365
Solibacillus silvestris StLB046	SSIL_2654	SSIL_2653
Sphingobium japonicum	SJA_C1-31180	SJA_C1-31190
Sanguibacter keddieii	Sked_00830	Sked_00820
Sanguibacter keddieii	Sked_33550	Sked_33540
Streptomyces leeuwenhoekii	sle_37960	sle_37970
Staphylococcus lugdunensis HKU09-01	SLGD_02162	SLGD_02161
Shewanella loihica	Shew_0618	Shew_0617
Syntrophothermus lipocalidus	Slip_0682	Slip_0683
Salinispira pacifica	L21SP2_0324	L21SP2_0323
Sideroxydans lithotrophicus	Slit_2261	Slit_2260
Streptomyces lividans	SLIV_17695	SLIV_17690
Streptomyces avermitilis	SAV_4077	SAV_4076
Sphingopyxis macrogoltabida	AN936_13765	AN936_13760
Sinorhizobium medicae	Smed_3394	Smed_3393
Sinorhizobium meliloti 1021	SMc02862	SMc02861
Stenotrophomonas maltophilia K279a	Smlt1705	Smlt1706
Stackebrandtia nassauensis	Snas_0240	Snas_0241
Sphingomonas sp. NIC1	A7E77_11925	A7E77_11920
Shewanella oneidensis	SO_3770	SO_3771
Staphylococcus pasteuri	STP1_1748	STP1_1749
Shewanella putrefaciens CN-32	Sputcn32_3096	Sputcn32_3097
Sphingopyxis fribergensis	SKP52_05270	SKP52_05275
Sphingomonas sp. MM-1	G432_09250	G432_09245
Sphingopyxis sp. 113P3	LH20_02830	LH20_02835
Shewanella pealeana	Spea_3487	Spea_3488
Spongiibacter sp. IMCC21906	IMCC21906_00280	IMCC21906_00281
Sporosarcina psychrophila	AZE41_02110	AZE41_02105
Streptomyces rapamycinicus	M271_22950	M271_22945
Selenomonas ruminantium	SELR_11620	SELR_11630
Streptosporangium roseum	Sros_0891	Sros_0892
Streptosporangium roseum	Sros_5568	Sros_5567
Streptosporangium roseum	Sros_7344	Sros_7343
Sphingomonas sanxanigenens	NX02_00340	NX02_00345
Staphylococcus schleiferi 1360-13	LH95_09970	LH95_09965
Staphylococcus pseudintermedius HKU10-03	SPSINT_0371	SPSINT_0372
Shewanella sediminis	Ssed_0369	Ssed_0370
Shewanella sediminis	Ssed_3843	Ssed_3844
Selenomonas sputigena	Selsp_0986	Selsp_0987

Staphylococcus simulans	AL483_08320	AL483_08315
Sulfolobus solfataricus P2	SSO2460	SSO2461
Staphylococcus saprophyticus	SSP2060	SSP2059
Streptomyces sp. SirexAA-E	SACTE_3577	SACTE_3578
Sphingomonas taxi	MC45_06045	MC45_06050
Stenotrophomonas sp. KCTC 12332	AXG53_10390	AXG53_10385
Sphingopyxis terrae	AOA14_12605	AOA14_12610
Sphaerobacter thermophilus	Sthe_0013	Sthe_0014
Sulfolobus tokodaii	STK_05880	STK_05890
Salinispora tropica	Strop_0282	Strop_0283
Sulfurimonas autotrophica	Saut_0947	Saut_0948
Stigmatella aurantiaca	STAUR_7027	STAUR_7026
Candidatus Solibacter usitatus	Acid_3187	Acid_3186
Streptomyces venezuelae	SVEN_3895	SVEN_3896
Streptomyces violaceusniger	Strvi_0601	Strvi_0602
Shewanella violacea	SVI_0532	SVI_0531
Streptomyces vietnamensis	SVTN_20360	SVTN_20365
Staphylococcus warneri	A284_10015	A284_10010
Shewanella woodyi	Swoo_0717	Swoo_0716
Sphingomonas wittichii	Swit_3606	Swit_3605
Syntrophomonas wolfei	Swol_1647	Swol_1646
Shewanella piezotolerans	swp_0233	swp_0231
Shewanella piezotolerans	swp_0813	swp_0812
Staphylococcus xylosus HKUOPL8	BE24_01055	BE24_01060
Thaumarchaeota archaeon MY3	NMY3_01011	NMY3_01012
Thermoplasma acidophilum	Ta0018	Ta0017
Thermosipho africanus	THA_217	THA_216
Thermosphaera aggregans	Tagg_1139	Tagg_1140
Candidatus Nitrosotenuis cloacae	SU86_003470	SU86_003465
Terribacillus aidingensis	GZ22_14420	GZ22_14415
Thermococcus barophilus	TERMP_01775	TERMP_01774
Thermobispora bispora	Tbis_0449	Tbis_0450
Thiolapillus brandeum	TBH_C1102	TBH_C1103
Thermodesulfobacterium commune	HL41_00045	HL41_00040
Thermobacillus composti	Theco_0668	Theco_0669
Thermomonospora curvata	Tcur_0588	Tcur_0589
Tepidanaerobacter acetatoxydans Re1	TepRe1_2058	TepRe1_2059
Thermococcus eurythermalis	TEU_01900	TEU_01905
Thermobifida fusca	Tfu_2745	Tfu_2744
Thermococcus gammatolerans	TGAM_0118	TGAM_0119
Thioalkalivibrio sulfidiphilus	Tgr7_2470	Tgr7_2469
Thermococcus guaymasensis	X802_04500	X802_04495
Thermococcus sp. AM4	TAM4_1353	TAM4_1264

Thermococcus sp. 4557	GQS_07640	GQS_07635
Thermogladius cellulolyticus	TCELL_0961	TCELL_0960
Tetragenococcus halophilus	TEH_14430	TEH_14440
Thermococcus cleftensis	CL1_1159	CL1_1158
Thermococcus paralvinellae	TES1_1766	TES1_1765
Thermococcus sp. 2319x1	ADU37_CDS08160	ADU37_CDS08170
Thermincola potens	TherJR_2053	TherJR_2054
Thermoanaerobacter kivui	TKV_c02810	TKV_c02800
Thioalkalivibrio sp. K90mix	TK90_0141	TK90_0142
Thermococcus kodakarensis	TK2060	TK2061
Pseudothermotoga lettingae	Tlet_0068	Tlet_0069
Pseudothermotoga lettingae	Tlet_0560	Tlet_0561
Thermococcus litoralis	OCC_08225	OCC_08230
Thermotoga maritima MSB8	TM0260	TM0261
Tatlockia micdadei	LMI_2756	LMI_2757
Thermosipho melanesiensis	Tmel_0788	Tmel_0789
Thermosipho melanesiensis	Tmel_1923	Tmel_1922
Thermoanaerobacter mathranii	Tmath_0367	Tmath_0366
Thermotoga neapolitana	CTN_0426	CTN_0425
Thermotoga naphthophila	Tnap_0890	Tnap_0891
Thermococcus nautili	BD01_1910	BD01_1909
Thermococcus onnurineus	TON_1551	TON_1552
Thermodesulfobacterium geofontis	TOPB45_0626	TOPB45_0625
Thermotoga petrophila	Tpet_0664	Tpet_0663
Turneriella parva	Turpa_2519	Turpa_2520
Trueperella pyogenes TP6375	CQ11_08975	CQ11_08970
Thermotoga sp. RQ2	TRQ2_0688	TRQ2_0687
Terriglobus saanensis	AciPR4_2206	AciPR4_2207
Terriglobus saanensis	AciPR4_2519	AciPR4_2520
Thermoanaerobacterium saccharolyticum	Tsac_0118	Tsac_0119
Thermococcus sibiricus	TSIB_1044	TSIB_1043
Candidatus Thioglobus singularis	W908_07925	W908_07920
Pseudothermotoga thermarum	Theth_0757	Theth_0756
Thermoanaerobacter tengcongensis	TTE0330	TTE0329
Thermosulfidibacter takaii	TST_0094	TST_0093
Thermobaculum terrenum	Tter_0128	Tter_0129
Teredinibacter turnerae	TERTU_0422	TERTU_0421
Thiocystis violascens	Thivi_0299	Thivi_0298
Thermoplasma volcanium	TVG0064561	TVG0063671
Thermoplasma volcanium	TVG0064976	TVG0063671
Thioalkalivibrio versutus	TVD_13400	TVD_13405
Thermoanaerobacterium xylanolyticum	Thexy_2002	Thexy_2003

Thermodesulfovibrio yellowstonii	THEYE_A1361	THEYE_A1362
Vibrio alginolyticus	N646_2578	N646_2577
Vibrio anguillarum 775	VAA_03758	VAA_03759
Variovorax paradoxus S110	Vapar_1469	Vapar_1470
Vibrio cholerae O1 biovar El Tor N16961	VC2441	VC2442
Vibrio coralliilyticus RE98	IX92_01860	IX92_01855
Vibrio sp. EJY3	VEJY3_01890	VEJY3_01885
Vibrio antiquarius	VEA_004528	VEA_004529
Aliivibrio fischeri ES114	VF_2239	VF_2240
Vibrio fluvialis	AL536_08070	AL536_08065
Vibrio furnissii	vfu_A00852	vfu_A00851
Vibrio campbellii	VIBHAR_00863	VIBHAR_00862
Vibrio harveyi	AL538_11445	AL538_11440
Verrucosispora maris	VAB18032_02180	VAB18032_02175
Vibrio mimicus	AL543_07005	AL543_07010
Vibrio nigripulchritudo	VIBNI_A0191	VIBNI_A0190
Candidatus Vesicomyosocius okutanii (Calyptogena okutanii)	COSY_0129	COSY_0130
Vibrio parahaemolyticus RIMD 2210633	VP0419	VP0418
Veillonella parvula	Vpar_1354	Vpar_1353
Aliivibrio salmonicida	VSAL_I2681	VSAL_I2682
Vibrio tasmaniensis	VS_0412	VS_0411
Vibrio tubiashii	IX91_12745	IX91_12750
Vibrio vulnificus CMCP6	VV1_0618	VV1_0619
Waddlia chondrophila	wcw_0291	wcw_0290
Wenzhouxiangella marina	WM2015_2146	WM2015_2147
Woeseia oceani	BA177_05605	BA177_05610
Weeksella virosa	Weevi_1367	Weevi_1368
Xanthomonas citri pv. citri 306	XAC1710	XAC1711
Xanthomonas axonopodis pv. citrumelo F1	XACM_1731	XACM_1732
Xanthomonas campestris pv. campestris ATCC 33913	XCC1690	XCC1691
Xylanimonas cellulosilytica	Xcel_0299	Xcel_0300
Xanthomonas oryzae pv. oryzae KACC 10331	XOO2975	XOO2974
Xanthomonas sacchari	SB85_20025	SB85_20020
Xanthomonas translucens	FD63_11050	FD63_11045
Zhongshania aliphaticivorans	AZF00_17625	AZF00_17630

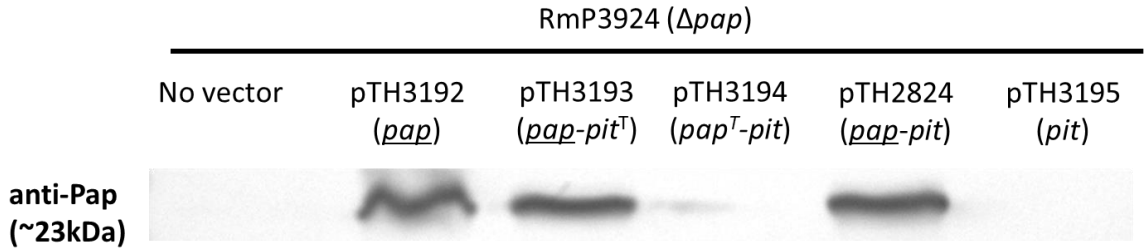


Appendix Figure 4 – anti-Rho1D4 western blot of Pap-Pit~Rho1D4 recombinant tag. Cells were lysed with buffer containing 2% (v/v) NP-40 and sonication prior to membrane fractional by centrifugation. Prior to SDS-PAGE, samples were mixed with 2X Laemmli buffer and incubated at room temperature for at least 30 minutes. There is a high amount of non-specific banding between the two samples and there does not appear to be any significantly different bands.

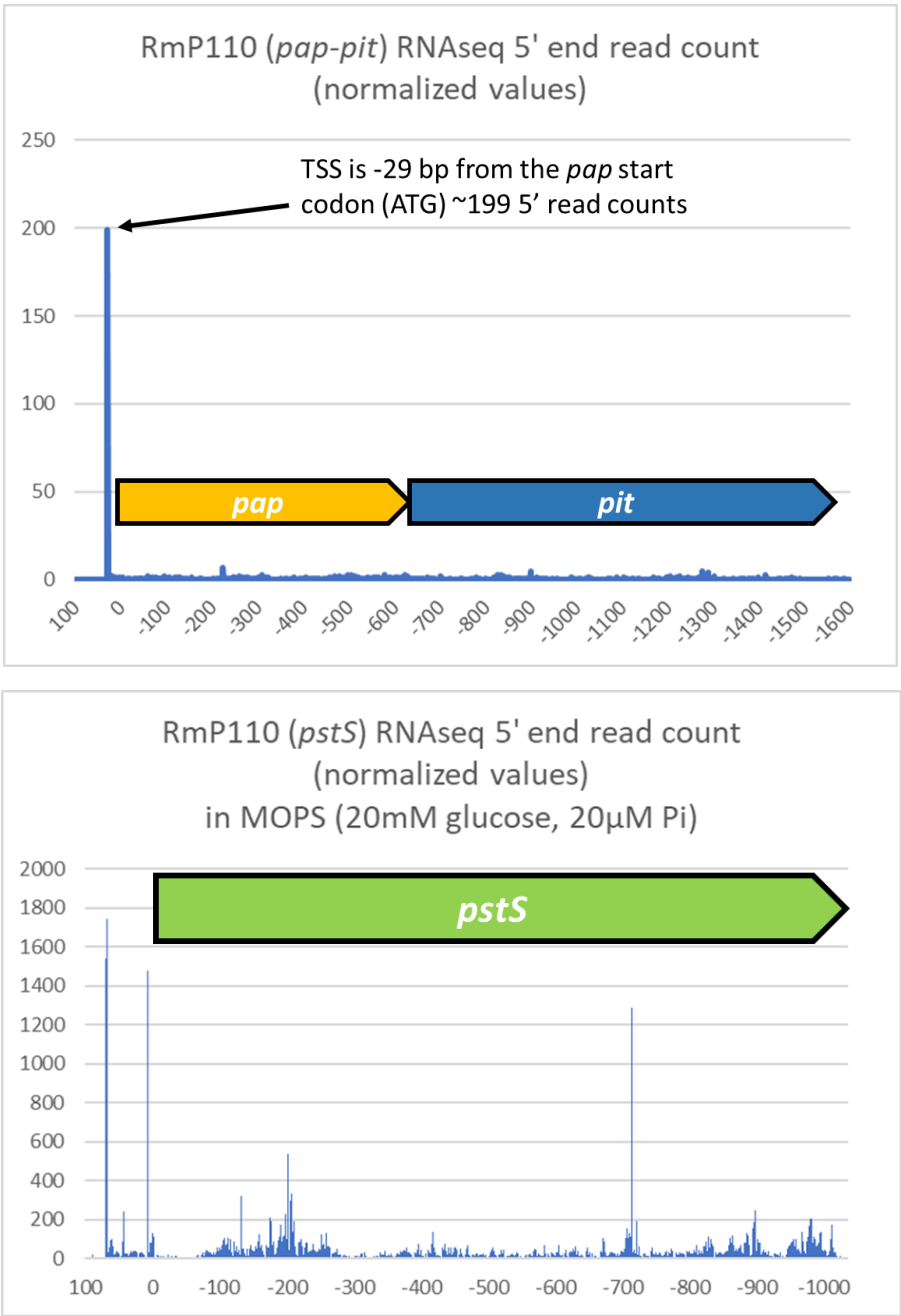
Appendix Table 5: B-galactosidase activity of BATCH two-hybrid constructs

Interaction	Media		
	LB	HiPi M9	LoPi M9
Pap-Pap	392 ± 12.9	124 ± 25.9	12 ± 5.1
Pap-Pit	14 ± 5.7	8 ± 5.9	21 ± 9.0
Pit-Pit	13 ± 2.2	16 ± 9.4	18 ± 5.1
Pap-Zip	8.1 ± 1.9	12 ± 5.5	9.0 ± 6.4
Pit-Zip	9.9 ± 4.7	6.1 ± 2.4	10 ± 7.1
Zip-Zip	212 ± 20.5	160 ± 18.8	182 ± 19.2

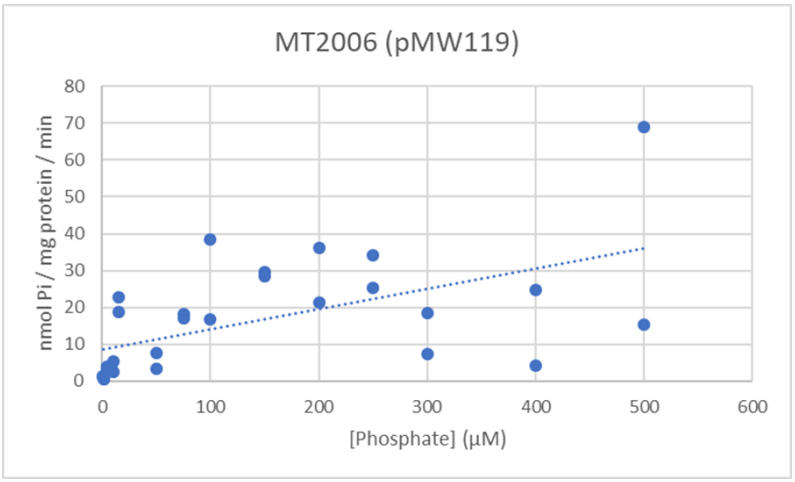
Appendix Table 5 – β-galactosidase activity from bacterial two-hybrid experiment of Pap-Pit interactions. Interactions between Pap-Pap, Pap-Pit, and Pit-Pit proteins were conducted using a bacterial two-hybrid system (BACTH), where each of the two proteins is cloned into a vector containing the T18 or T25 fragments of the *B. subtilis* adenylate cyclase enzyme which generate an N- or C-terminal translational fusion. The table above shows the mean highest observed β-galactosidase activity observed for the interactions. A leucine zipper fusion with the T18 and T25 translation fusions (Zip-Zip) was used as a positive control for a protein-protein interaction, whereas a Pap- or Pit-Zip interaction was used as a negative control. The values shown are in Miller Units ± Standard Deviation.

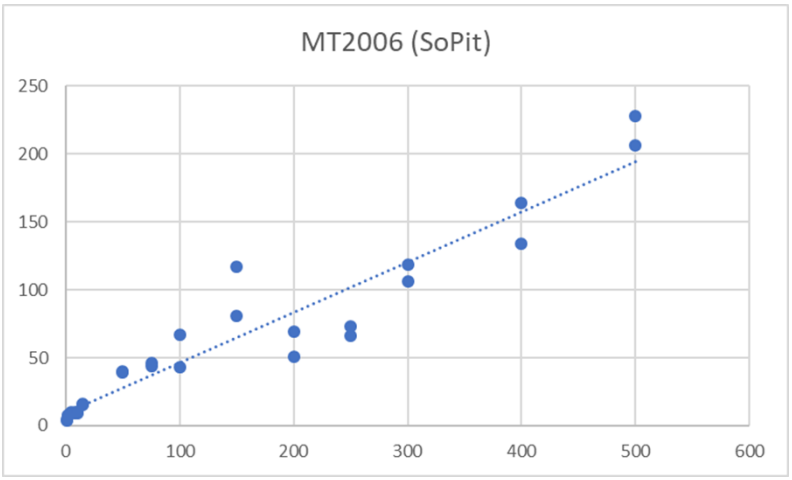


Appendix Figure 6 – Anti-Pap western blot of the RmP3924 (Δ*pap*) recombinants. Detection of the Pap protein using the anti-Pap antisera demonstrated that the Pap protein was present in all scenarios where the *pap* gene was present, which was as expected.

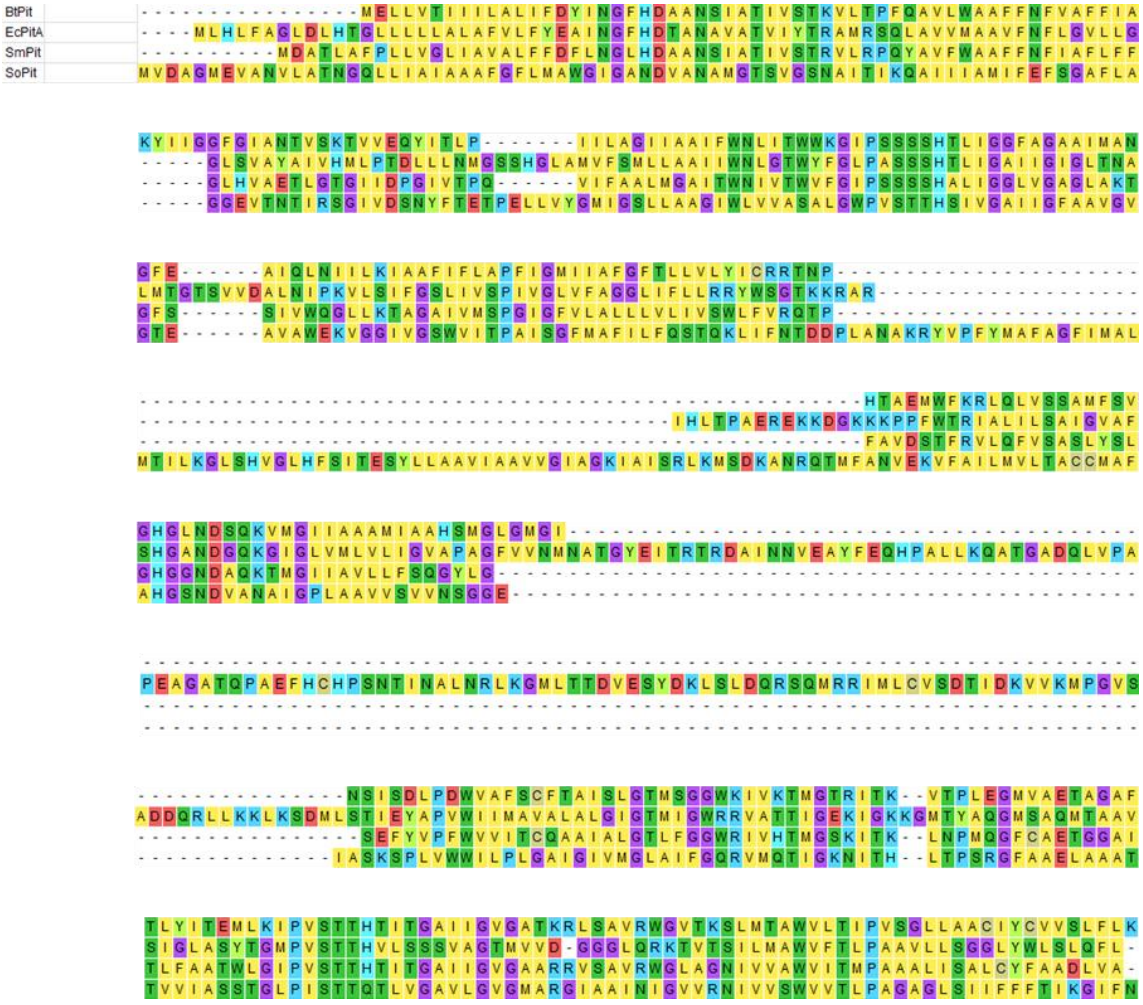


Appendix Figure 7 – RNAseq data shows that *S. meliloti pap-pit* is expressed at low levels. RNAseq 5' end read normalized counts of wildtype *S. meliloti* RmP110 *pap-pit* are expressed at moderate levels in MOPS minimal medium with high phosphate (2 mM Pi) with approximately 200 reads. Whereas the *pst* operon is highly induced when cells are grown in MOPS minimal medium with low phosphate (20 µM Pi) with read counts at approximately ~1700. This indicates that comparatively, the *pst* operon is transcribed 8 times than the *pap-pit* operon in the respective media.

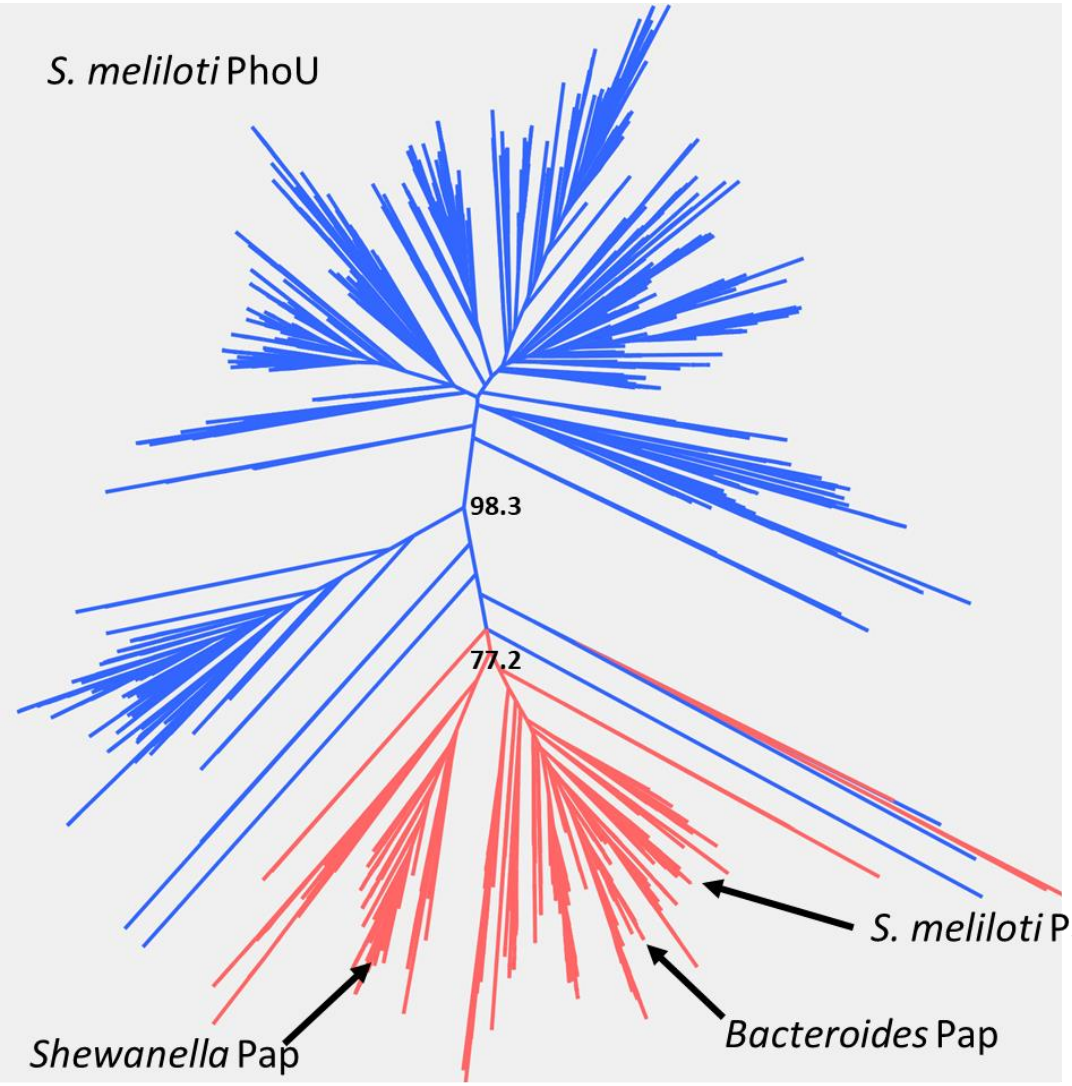




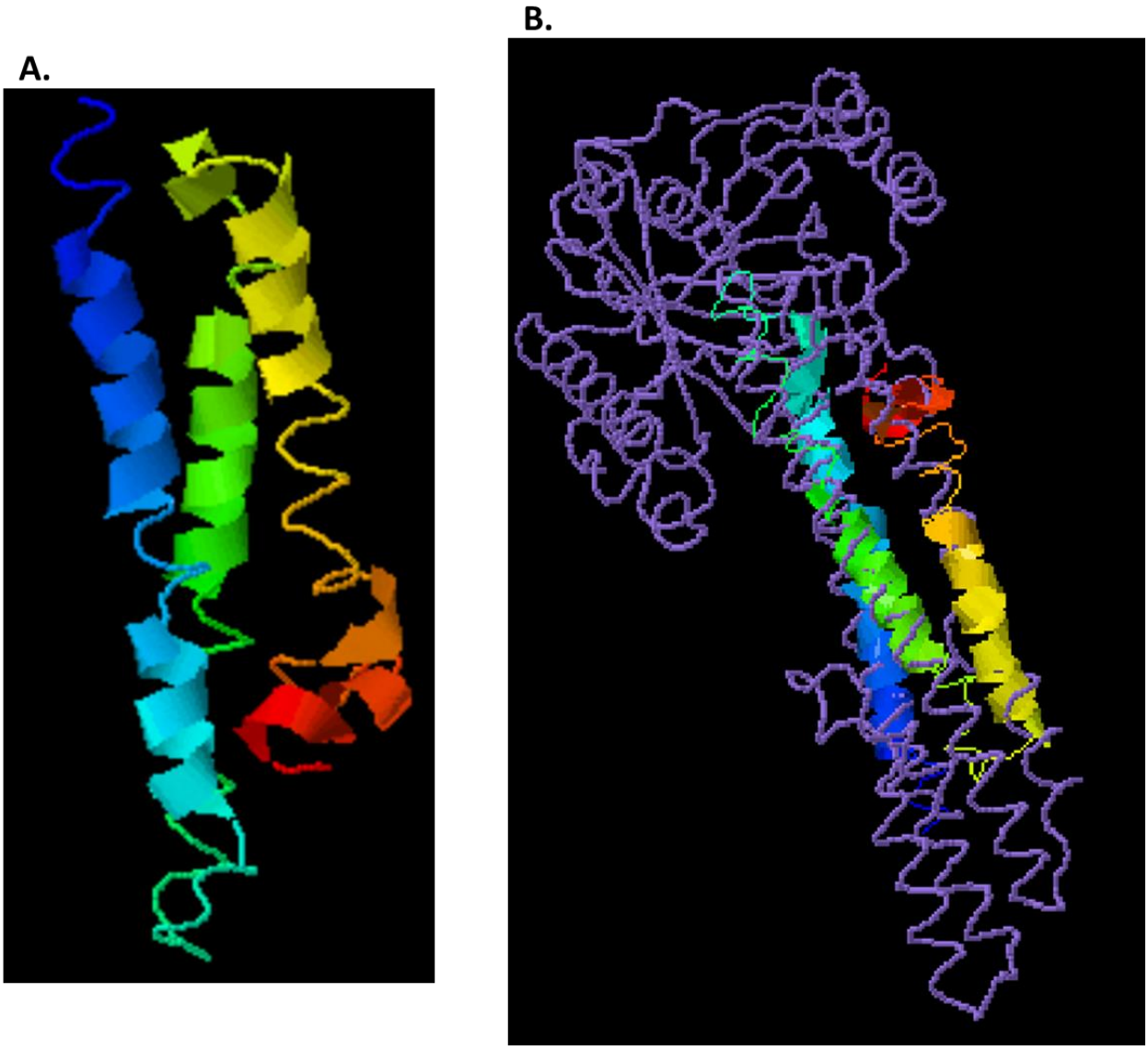
Appendix Figure 8 – Kinetics of Pi uptake for *E. coli* MT2006 (pMW119). The velocity of Pi uptake for *E. coli* MT2006 (pMW119) was measured at concentrations varying between 0.25 µM and 500 µM Pi in order to determine if there were background levels of PI uptake in *E. coli* MT2006 due to transporters active only at higher concentrations. However, the velocity of Pi uptake for this strain was not observed to be linear, nor was the velocity consistent (variable between the two replicates conducted). Additionally, although there was some observable Pi accumulation within the cells, this accumulation appeared to increase and decrease with higher concentrations of Pi.



Appendix Figure 9 – Alignment of the Pit orthologues in this study (SmPit, BtPit, SoPit, and EcPitA). Alignment of the four proteins was conducted using ClustalW and coloured according to amino acid type. The transmembrane helices were predicted using TMHMM and are boxed in red. PHO4 motifs are marked with a black line. Interestingly, we observe two major gap regions between these proteins. First, there is a ~50 a.a. insertion between the two PHO4 domains found in SoPit but not any of the other proteins. Secondly, there is a 124 a.a. insertion within the C-terminal PHO4 domain of the EcPitA protein that is not present in the other proteins. These regions of interest may explain the functional differences observed between the Pit proteins.



Appendix Figure 10 – Maximum Likelihood Phylogeny of Pap and PhoU proteins. Pap (red) and PhoU proteins (blue) were observed to form two distinct clades (bootstrap value of ~98%), with few mixing between the two groups. This indicates that Pap and PhoU proteins are distinct proteins, where the two proteins diverged some point in evolutionary history. Additionally, similar to the Pit protein phylogeny, *S. meliloti* and *Bacteroides* Pap are more closely related compared to *Shewanella* Pap.



Appendix Figure 11 – Structural prediction of the EcPitA-loop region compared with the SPX domain of *S. cerevisiae* Vtc4 (inorganic polyphosphate polymerase). A. I-TASSER structural prediction of the EcPitA-loop region (~124 a.a.) resulted in a structure that contained 3 parallel α-helices however with poor accuracy (TM-score = 0.42 ± 0.14; RMSD 9. 9± 4.6 Å), as there are no crystal structures in PDB which share sequence similarity with the EcPitA protein. **B.** TM-Align identified THE SPX-TMM domain of the *S. cerevisiae* inorganic polyphosphate polymerase protein Vtc4 as a protein with a similar structure (RMSD 3.42 Å). Intriguingly, the helices of the putative EcPitA loop structure matched very well with the parallel helices of the SPX domain.