Understanding the Catalytic Cycle of Membrane Pyrophosphatases Through Structural and Functional Studies

by

Craig Wilkinson

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

School of Biomedical Sciences
&
Astbury Centre for Structural Molecular Biology

October 2017

Intellectual property and Publication Statement

The candidate confirms that the work submitted is his/her/their own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly stated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others. Chapters 4, 5 and 6 are in part based on work that was submitted in jointly authored publications. The publications in question are:

(1) "Li, K. M.*, C. Wilkinson*, J. Kellosalo, J. Y. Tsai, T. Kajander, L. J. Jeuken, Y. J. Sun and A. Goldman (2016). "Membrane pyrophosphatases from *Thermotoga maritima* and *Vigna radiata* suggest a conserved coupling mechanism." Nat Commun **7**: 13596"

In which I am a joint first author. For this paper, I solved two new crystal structures of TmPPase using data collected by Dr Juho Kellosalo and Dr Tommi Kajander from protein expressed and purified by Dr Juho Kellosalo. Our collaborators from Taiwan carried out the expression, purification and crystallisation of the VrPPase and carried out the structural refinement of their protein. They also supplied me with the VrPPase protein that was used in the SURFE2R experiments. I carried out these experiments in the lab of Dr Lars Jeuken and am responsible for the data collection and analysis. I carried out the structural analysis and updating of the catalytic cycle model with input from all authors.

(2) Shah, N. R., C. Wilkinson, S. P. Harborne, A. Turku, K. M. Li, Y. J. Sun, S. Harris and A. Goldman (2017). "Insights into the mechanism of membrane pyrophosphatases by combining experiment and computer simulation." Struct Dyn **4**(3): 032105.

Only the additional SURFE^{2}R experiments from this publication were used in my thesis, I carried out all referenced experiments myself in the lab of Dr Lars Jeuken and am responsible for setting up the experiments, data collection and subsequent analyses.

Acknowledgements

I would first like to thank my Supervisors, Prof. Adrian Goldman and Dr. Stephen Muench, for the opportunity to carry out this research and the mountains of support given throughout this process. Particularly for constantly challenging me to be a better scientist and person, it has truly been a pleasure to work on this project.

Thanks also to my predecessor, Dr. Juho Kellosalo, for his constant advice and generosity on this project. To Prof. Peter Henderson, for his comments and feedback in assessing my work over the course of this project and to David Sharples for his vast array of technical support.

Thank you to Dr. Lars Jeuken and Matthias Gantner for their help and support in getting up and running with the Nanion SURFE^{2}R technique; and to the lab of Prof. Yuh-Ju Sun, including Dr. Kun-Mou Li and Dr Jia-Yin Tsai, for providing protein and knowledge for the SURFE^{2}R experiments and the highly valuable collaboration on work in the M-PPase field, I am truly proud of what we have achieved.

My thanks go to everyone from the Goldman lab (Leeds and Helsinki, past and present) and Astbury level 6 for making my time here in Leeds a productive and thoroughly enjoyable experience. Particular thanks go to: Dr. Nita Shah, for her help in getting me set up in the lab along with her optimism and willingness to always answer my questions. To Dr. Maren Thomsen for her overwhelming support in all things crystallography and beyond, I certainly wouldn't have made it this far without her advice and guidance. To Dr. Steven Harborne for his guidance in expressing and purifying membrane proteins and for the constant IT support at the expense of his own work time. Thanks also to Jamie Henderson, Brendan Farrell and Dr. Julie Heggelund for their constant support both in and out of the lab, it is much appreciated.

Thanks also to everyone else, past and present, including (in no particular order): Yurui Ji, Robert Kolodziejczyk, Jack Wright, Jessica Boakes, Jannik Strauss, Claudia Stohrer, Simon Davies, Rhys Bryant, Daniel Fludgate, Amelia Lesiuk, Sophie Hesketh, Ana Pavic, Francis Totanes, Alex Holmes and all my friends and family for their unending support and the unenviable task of keeping me sane, answering a never-ending source of

questions, and listening to my increasing brilliant, albeit utterly irrelevant, ideas whilst procrastinating!

A final thanks to everyone listed, and beyond, for their role in bringing me out of my shell, helping me find my voice and for making me a better researcher and human being. This has truly been a life-changing experience for me, one that I will never forget.

Abstract

Membrane pyrophosphatases (M-PPases) couple pyrophosphate hydrolysis to the translocation of sodium ions/protons, using the resulting ion gradients to drive abiotic stress resistance and in the infectivity of protozoan parasites. I have solved two M-PPase structures in different catalytic states, combining these with previous structures to update the model of the catalytic cycle of M-PPases. These new structures confirm previous findings that substrate binding breaks interactions between $K^{12.50}$ and $D^{6.43}$ due to motion of helix 12, leading to a rearrangement of helix 6 and priming the enzyme for hydrolysis. Previously this information was only visible between the structures of two-distinct M-PPases, a H^+-PPase and Na^+-PPase. The current structures allow for comparisons to be made between structures of the same type of M-PPase. Electrometric data was acquired using the Nanion SURFE^{2}R technique, which showed a proton-pumping signal was generated by the non-hydrolysable inhibitor, imidodiphosphate. This provided sufficient information to update the model of the complete catalytic cycle, favouring the hypothesised Binding change mechanism, in which substrate binding induces a series of conformational changes during which ion pumping occurs first, followed by substrate hydrolysis.

Additionally, crystal optimisation techniques improved the resolution of the *Pyrobaculum aerophilum* M-PPase structure to 3.8, providing an overview of the K^+-independent M-PPase. The hydrolytic centre and ion gate regions showed similar coordination to previous structures, with differences seen in the conformation of several outer ring helices, potentially linked to K^+-independence. I also carried out mutational studies investigating $K^{12.46}$ and $T^{12.49}$, both involved in K^+-independence and found that both mutations were required to generate a K^+-dependent variant of PaPPase. Overall, this information has improved our understanding of the structure and function of the membrane pyrophosphatases, providing a basis for drug-design programmes targeting protozoan parasites, to which the membrane pyrophosphatases are a vital part of growth and infectivity.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ANS: 8-anilino 1-naptholene sulphonate

ATP: Adenosine Triphosphate

BgG: Bovine gamma Globulin

BSA: Bovine Serum Albumin

CCCP: Carbonyl cyanide m-chlorophenyl hydrazone

CMC: Critical Micellar concentration

CV: Column Volume

CYGLU-3: 3-Cyclohexyl-1-Propyl-β-D-Glucoside

CYMAL-7: 7-Cyclohexyl-1-Heptyl-β-D-Maltoside

DDM: n-Dodecyl−β−D-Maltopyranoside

DDTM: n-Dodecyl−β−D-Thiomaltopyranoside

DLS: Diamond Light Source

DM: n-Decyl−β−D-maltopyranoside

DMSO: Dimethyl Sulphoxide

DNA: Deoxyribo Nucleic Acid

DTT: Dithiothreitol

EDTA: Ethylenediaminetetraacetic acid

EGTA: Ethylene glycol-bis(2-aminoethylether)-N.N.N',N'-tetraacetic acid

ESRF: European Synchrotron Radiation facility

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

His: Histidine

HTG: n-Heptyl−β−D-Thioglucopyranoside

IDP: Imidodiphosphate

LMNG: Lauryl maltose neopentyl glycol

M-PPase: Membrane Pyrophosphatase

MEDP: Methlyenediphosphate

MES: 2-(N-Morpholino)ethanesulphonic acid

MOPS: 3-(N-morpholino)propanesulfonic acid

NCS: Non-Crystal symmetry

NG: n-Nonyl−β−D-Glucopyranoside

NTA: Nitrilotiacetic Acid

NTM: n-Nonyl−β−D-Thiomaltopyranoside

DTM: n-Decyl−β−D-Thiomaltopyranoside

OGNG: octyl glucose neopentyl glycol

OTM: n-Octyl−β−D-Thiomaltopyranoside

Pa: *Pyrobaculum aerophilum*

PAGE: Polyacrylamide gel electrophoresis

PCR: Polymerase Chain Reaction

PDB: Protein Data Bank

PEG: Polyethlyene Glycol

P_i: Inorganic phosphate

PMSF: Phenylmethylsulphonyl fluoride

PPase: Pyrophosphatase

PP_i: Inorganic pyrophosphate

S-PPase: Soluble Pyrophosphatase

SCD-Leu: Synthetic Complete Dropout (minus leucine)

SDS: Sodium dodecyl sulphate

TBS: Tris Buffer Saline

Tm: *Thermotoga maritima*

UM: n-Undecyl−β−D-Maltopyranoside

UM: n-Undecyl−β−D-Maltopyranoside

UTM: n-Undecyl−β−D-Thiomaltopyranoside

Vr: *Vigna radiata*

YPD: Yeast extract Peptone Dextrose

Chapter 1: Introduction and Background on Pyrophosphatases

1.1 Types of Pyrophosphatase

Inorganic pyrophosphatases (PPases) are proteins found in all forms of life. They catalyse the hydrolysis of pyrophosphate (PP_i) into two phosphate (P_i) molecules (Kornberg, 1962; Moe & Butler, 1972). PP_i is a product of many vital anabolic processes such as nucleotide, lipid and protein synthesis. By removing PP_i, PPases provide a thermodynamic driving force that pushes these metabolic processes in the biosynthetic direction by rapidly reducing the PP_i pool in the cytoplasm (Kornberg, 1962). In this way, they prevent a toxic build-up of PP_i that would inhibit essential biosynthetic pathways.

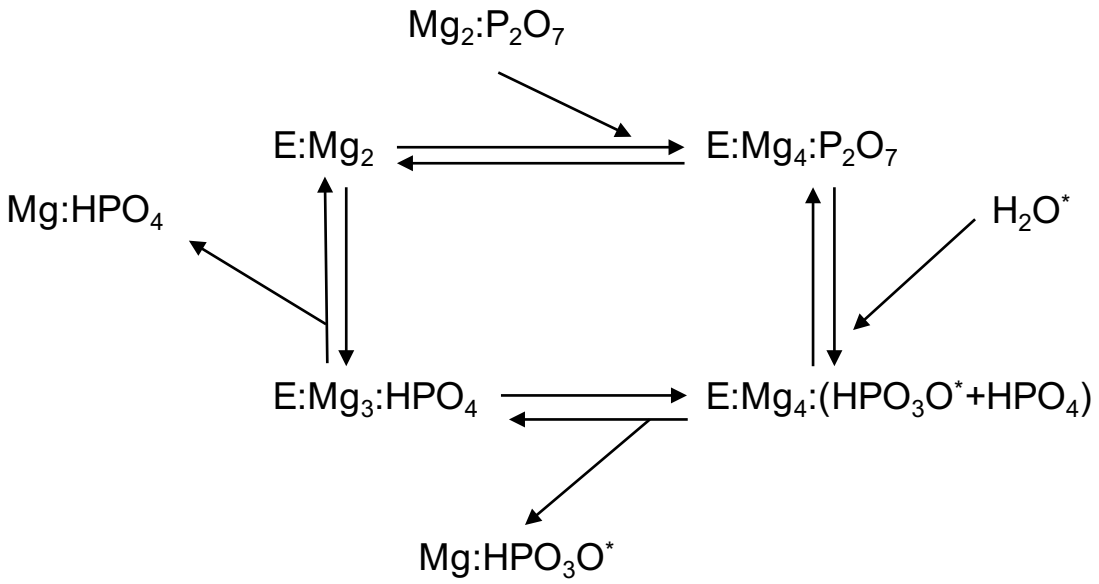
There are two main groups of PPases: soluble PPases (S-PPases) and membrane PPases (M-PPases). The S-PPases are found in all known living organisms and are essential for life. Decreased S-PPase expression in gene disruption experiments led to an increase in intracellular PP_i concentration and reduced cellular growth rates (Chen *et al.*, 1990) and gene knockout experiments resulted in non-viable cells (Lundin *et al.*, 1991).

In contrast, M-PPases are found only in plants, protozoan parasites and some species of bacteria and archaea. They hydrolyse PP_i into P_i at a significantly slower rate (20-200x slower than S-PPases) (Hsu *et al.*, 2009; Parfenyev *et al.*, 2001) but couple this to the translocation of cations such as sodium ions or protons across the membrane in order to establish an ion gradient (Malinen *et al.*, 2007). Pyrophosphate synthesis (P_{i2} to PP_i) has also been observed, generally occurs at much slower rates and is only capable of generating an equilibrium rather than performing active transport (Belogurov *et al.*, 2005).

1.1.1 Type I Soluble Pyrophosphatases

S-PPases were first discovered in 1928 from extracts of mammalian tissue culture that had a detectable pyrophosphatase activity (Kay, 1928) and later from preparations of yeast fermentations, where the addition of pyrophosphate led to an increase in phosphate concentration (Boyland, 1930). The first extracted forms of the enzymes were obtained by Kunitz (1951) who was able to repeatedly isolate crystalline forms of the protein for use in various studies.

Hydrolysis of the substrate was found to occur due to direct nucleophilic attack from a water molecule, transferring the oxygen from a water molecule to the PP_i, resulting in two phosphate molecules (Cohn, 1958; Kunitz, 1951). An electrophilic phosphate contains the oxygen atom from the water molecule and the leaving group phosphate contains the oxygen atom from PP_i. It was noted that metal ions were essential for substrate binding and hydrolysis, negating the charge of the pyrophosphate (Cohn, 1958), with further studies revealing that two metal ions could bind to the native S-PPase in the absence of substrate (Rapoport *et al.*, 1973) along with a third in the presence of P_i (Cooperman *et al.*, 1981; Springs *et al.*, 1981). Various metal ions were tested but only four showed any activity, which in descending order of activity are: magnesium, zinc, cobalt and manganese ions (Butler & Sperow, 1977; Höhne & Heitmann, 1974; Janson *et al.*, 1979). A simplified scheme of the hydrolysis mechanism of S-PPases was generated (Scheme 1).



Scheme 1: Simplified reaction scheme for Pyrophosphatases, showing the hydrolysis of PP_i and sequential product release.

Experiments were performed measuring the effect of phenylglyoxal (a compound that specifically modifies arginine), on the active site residues of S-PPases. These found that the active site contained an essential arginine (Bond *et al.*, 1980; Cooperman & Chiu, 1973). Other inactivation and protection experiments (wherein the protein activity is measured following treatment with compounds with different chemical properties) were also carried out on the S-PPases. Compounds, such as N-acetlyimidazole (Negi *et al.*,

1972) and iodoacetamide (Yano *et al.*, 1973) showed that the active site might contain aspartate/glutamate, tryptophan and methionine residues (Cooperman & Chiu, 1973).

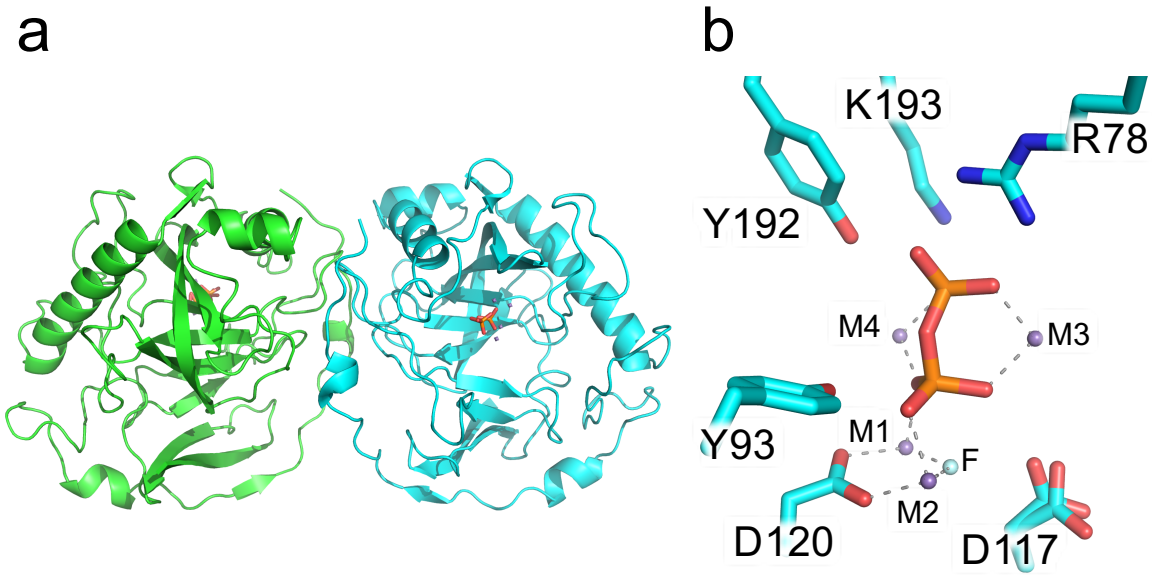


Figure 1.1: Structural overview of the Type-I Soluble Pyrophosphatases.
View of the type-I S-PPase from S. cerevisiae showing (a) dimer of 1.1 Å P_{i2}-bound structure (PDB ID: 1E9G) showing monomers A (green) and B (cyan) and (b) close up of the active site coordination of fluoride inhibited PP_i-bound structure (PDB ID: 1E6A), where Fluoride occupies position of the nucleophilic water.

In 1974 it was shown that crystals of the yeast S-PPase would be suitable for x-ray crystallography experiments (Bunick *et al.*, 1974). In 1978 the first structure was solved of a yeast S-PPase (Bunick *et al.*, 1978) followed independently by a structure from Harutyunyan, *et al* in 1981, which was not published (data unavailable). Structures of the *E. coli* S-PPase were solved in 1994 and 1995 (Kankare *et al.*, 1994) (Salminen *et al.*, 1995) with additional structures of the yeast S-PPase in 1996 (Harutyunyan *et al.*, 1996), including high resolution structures at 2 Å (Heikinheimo *et al.*, 1996) and 1.1 Å (Heikinheimo *et al.*, 2001). The overall structure consists of multiple helices and strands with the highly conserved core of the protein forming an "OB-fold" domain, consisting of 5 beta strands forming a closed barrel with the active site present on one face of the barrel (Heikinheimo *et al.*, 1996) (Figure 1.1a). The active site consists of several predicted metal ions coordinating the substrate and being vital for activity (Figure 1.1b). In addition, many of the predicted residues can be found in the active site, including several aspartate, glutamate, lysine, tyrosine and arginine residues, but no methionine was present.

1.1.2 Type II Soluble Pyrophosphatases

A second type of S-PPase was discovered, originally identified in *Bacillus subtilis* (Shintani *et al.*, 1998) and was found to be sequentially distinct from the type-I S-PPases, with the structure being later solved in *Streptococcus mutans* (Merckel *et al.*, 2001) and *B. subtilis* (Ahn *et al.*, 2001). These are structurally distinct but functionally analogous to the type-I S-PPases and are part of the DHH family of phosphoesterases, named after the conserved Asp-His-His motif (Aravind & Koonin, 1998). They consist of two distinct domains, the N-terminal domain containing a 5-strand β-sheet, upon which the active site is located facing the domain interface, along with seven helices (Figure 1.2a). The C-terminal domain also consists of a 5-strand mixed β-sheet and several helices. This global structure and the related folds are completely different from those forming the type-I S-PPases, but they still perform the same function, the hydrolysis of PP_i to P_{i2}. The active site is the most conserved region of the type-II S-PPases and contains 4 aspartate and 2 histidine residues that are completely conserved (Figure 1.2b). These proteins also bind 3 or 4 metal ions (Parfenyev *et al.*, 2001), with a propensity for manganese ions over magnesium. The structure shows several water molecules, including one believed to be the nucleophile, bound at the active site coordinated by three manganese ions and the substrate/product, which are in turn coordinated by a histidine residue within the active site (Merckel *et al.*, 2001).

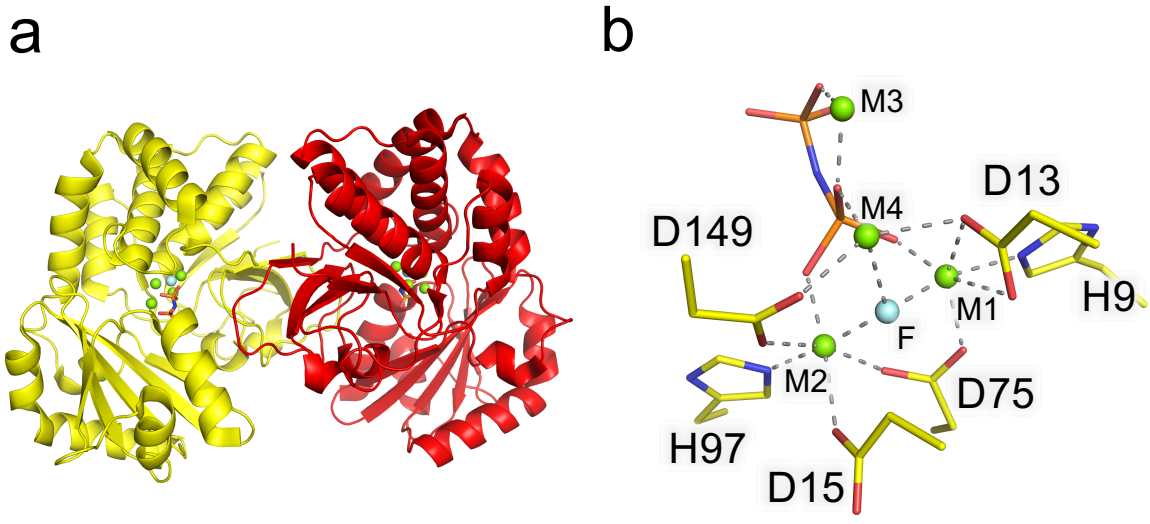


Figure 1.2: Structural Overview of the Type-II Soluble Pyrophosphatases.
View of the 1.75 Å fluoride inhibited IDP-bound structure of type-II S-PPase from Bacillus subtilis (PDB ID: 1E6A) showing (a) dimer composition with monomers A (green) and B (cyan) and (b) close up of the active site coordination between residues, metal ions and IDP, where Fluoride occupies position of the nucleophilic water.

The superimposition of this structure with the type-I yeast S-PPase showed a high degree of similarity in the coordination of the metal ions, substrate and water molecules, leading the authors to conclude that the mechanism of action was likely the same as in type-I S-PPases, despite the differences in the global structure (Merckel *et al.*, 2001). However, later studies found evidence to suggest that the two types functioned in completely different ways. They have a similar rate of activity in the presence of Mg^{2+}, but type-II S-PPases have a much higher rate of activity in the presence of Mn^{2+} compared with the Type-I S-PPases (Zyryanov *et al.*, 2004), likely due to the histidine residues in the active site, which preferentially bind Mn^{2+} over Mg^{2+}. The rate of catalysis was around 10-fold higher in type-II PPases (~3,000 s^{-1}) compared with type-I S-PPases (~200 s^{-1}) (Parfenyev *et al.*, 2001).

1.1.3 Type III Soluble Pyrophosphatases

Two distinct S-PPases have also been discovered in the thermophilic archae: *Thermococcus onnurineus* (Lee *et al.*, 2009) and in *Bacteroides thetaiotaomicron* BT2127 (Huang *et al.*, 2011). Both of these enzymes have been grouped into a third category of S-PPases, termed type-III S-PPases. Unlike the other types, these enzymes are Ni^{2+}-dependent (Lee *et al.*, 2009) and have so far only been located in a few bacterial species. Structurally these proteins resemble other members of the haloalanoic acid dehalogenase superfamily (HADSF), to which this class of proteins appears to belong (Huang *et al.*, 2011). These proteins have an α/β Rossmann-like fold composed of a six-stranded β-sheet surrounded by six helices and with a four-helix cap domain (Huang *et al.*, 2011) (Figure 1.3a). The active site differs greatly from the other S-PPases, with an aspartate nucleophile attacking the pyrophosphate (Figure 1.3b). The remaining catalytic residues are located on a series of 4 loops that form the active site. The main catalytic mechanism therefore more closely resembles other members of the HADSF than that of other pyrophosphatases.

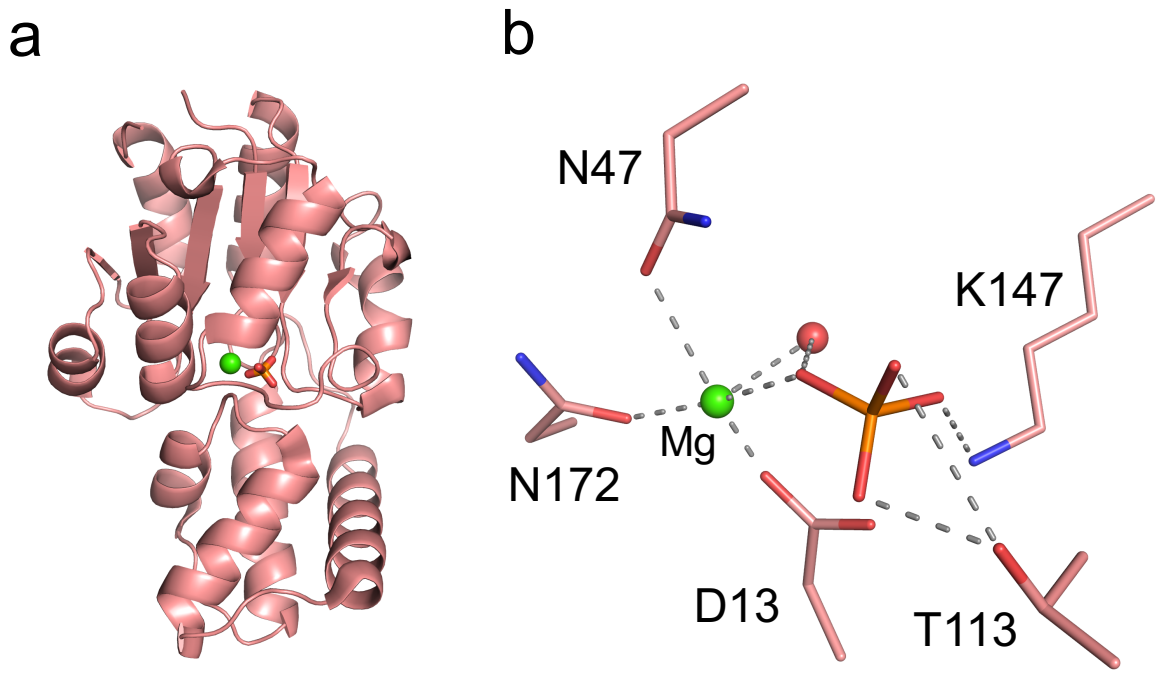


Figure 1.3: Structural Overview of the Type-III Soluble Pyrophosphatases
View of the 1.6 Å phosphate-bound E47N mutant of Type-III S-PPase from Bacteroides thetaiotaomicron (PDB ID: 3QYP) showing (a) monomer showing bound phosphate and magnesium ion and (b) close up of the active site coordination between residues, metal ions and P_i.

All types of S-PPase, despite the structural differences, have identical functions in recycling intracellular PP_i in order to maintain the equilibrium of biochemical reactions and allow biosynthetic reactions to continue. In addition to the S-PPases, a distinct class of PPases was also discovered, the membrane pyrophosphatases (M-PPases).

1.1.4 Membrane Pyrophosphatases

The M-PPases have additional function in cells compared to S-PPases, coupling the pyrophosphatase activity with the active membrane translocation of protons (Moyle *et al.*, 1972) and/or sodium ions (Malinen *et al.*, 2007) (Figure 1.4). The catalytic rate of these enzymes is much slower than any of the S-PPases, 10-fold slower than the type-I S-PPases and ~100-fold slower than the type-II S-PPases (Hsu *et al.*, 2009; Parfenyev *et al.*, 2001). However, their ability to pump cations contributes to cell homeostasis by establishing membrane potentials and is particularly important during periods of environmental stress (García-Contreras *et al.*, 2004; López-Marqués *et al.*, 2004; Nyrén & Strid, 1991). This confers a significant advantage to organisms frequently subjected to these stresses, as PP_i released as a waste product during primary anabolic processes can

be used to generate ion gradients (Kornberg, 1962). By activating these enzymes, vital homeostatic mechanisms (such as pH and osmoregulation) can be maintained using these gradients, even in ATP-deficient environments such as in plants during drought stress (Venancio *et al.*, 2014).

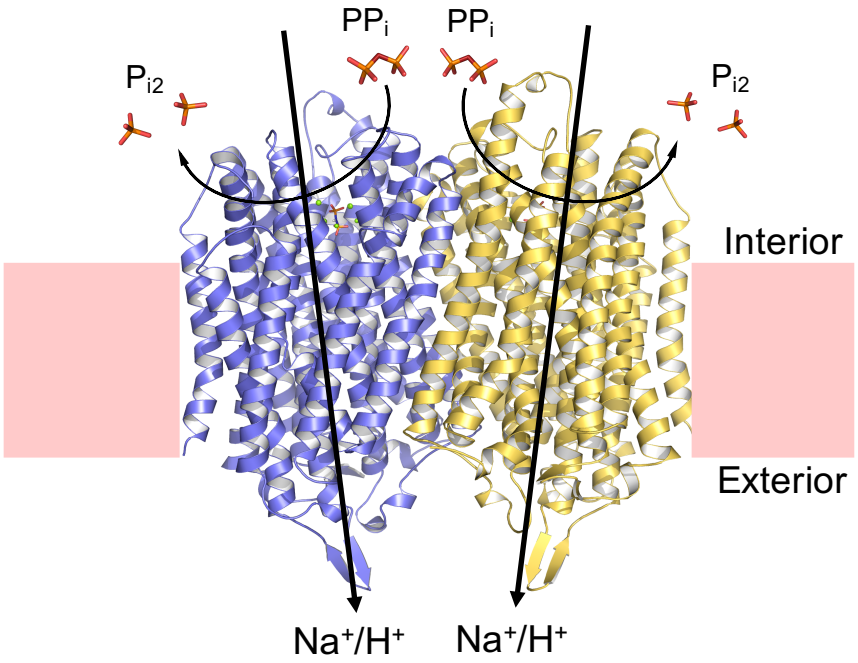


Figure 1.4: Generalised Overview of Membrane Pyrophosphatase Function.
Cartoon representation of a generalised M-PPase dimer showing the PP_i to P_{i2} hydrolytic activity and Na^+ and/or H^+ translocation function.

M-PPases are primary ion pumps, meaning that the energy utilised to pump ions across the membrane is obtained directly from the hydrolysis of PP_i, in this case occurring in the same active site channel (Kim *et al.*, 1994). This differs from secondary pumps where this energy would be from ion or proton gradients established by other proteins or domains. Having all the required machinery in the same subunit provides a greater degree of control over enzyme function, ensuring that all of the energy generated from pyrophosphate hydrolysis goes towards establishing the ion gradient.

M-PPases were first identified in *Rhodospirillum rubrum*, a photosynthesising bacteria found in water sources following studies performed on chromatophores in which PP_i was produced during photophosphorylation (Baltscheffsky *et al.*, 1966). This process was later attributed to a single enzyme, the M-PPase when translocation of protons across the membrane was observed in the presence of PP_i and was not attributable to other enzymes,

such as H^+-ATPases (Moyle *et al.*, 1972). M-PPases have so far been found in a selection of different species, including plants (Bucke, 1970; Karlsson, 1975), protozoan parasites (Scott *et al.*, 1998), some bacteria (Baltscheffsky *et al.*, 1966) and archaea (Drozdowicz *et al.*, 1999).

The localisation of this protein differs between species. In bacteria and archaea they are generally located in the plasma membrane or in organelle-like structures (Seufferheld *et al.*, 2003), whereas in plants and protozoan parasites they are predominantly found in the vacuole (Rea *et al.*, 1992) and acidocalcisome organelle (Scott *et al.*, 1998), respectively. Thus far, no similar enzymes have been found in other high-order Eukaryotic species. Since their discovery, work has focussed on understanding the mechanism of action and respective roles of these proteins in specific organisms.

1.2 Classification of M-PPases

1.2.1 Types of M-PPase

The primary categorisation of M-PPases is based on cation specificity, with some M-PPases exclusively pumping protons (H^+-PPases), others exclusively pumping sodium ions (Na^+-PPases) and a third group that is capable of pumping both sodium ion and protons (Na^+/H^+-PPases) (Luoto *et al.*, 2013) (Table 1.1). These types of M-PPase can be further sub-divided on the requirement of K^+ for activity. All Na^+-PPases and the Na^+/H^+-PPases are K^+-dependent, along with some H^+-PPases (K^+-dependent H^+-PPases) (Karlsson, 1975), but there are a distinct group of H^+-PPases that are insensitive to K^+, showing full activity even in the absence of K^+ (K^+-independent H^+-PPases) $^+$ (Drozdowicz *et al.*, 1999). Moreover, the Na^+/H^+-PPases can also be subdivided, as some can only pump protons in low sodium concentrations (Na^+-dependent Na^+/H^+-PPases) and others can pump both sodium ions and protons regardless of sodium concentration (True Na^+/H^+-PPases) (Nordbo *et al.*, 2016) (Table 1.1).

Table 1.1: Types of Membrane Pyrophosphatase
List of all known types of membrane pyrophosphatase, showing name and characteristics of each type

Type	Pumping substrate	K^+-sensitivity	Examples
Na^+-PPase	Na^+	Yes	*Thermotoga maritima*
K^+-dependent H^+-PPase	H^+	Yes	*Vigna radiata, Plasmodium falciparum* VP1
K^+-independent H^+-PPase	H^+	No	*Pyrobaculum aerophilum, Plasmodium falciparum* VP2
Na^+-dependent Na^+/H^+-PPase	Na^+ and H^+ (Only H^+ at low $[Na^+]$)	Yes	*Clostridium leptum*
True Na^+/H^+-PPase	Na^+ and H^+	Yes	*Bacteroides vulgatus*

M-PPase capable of pumping Na^+ (either exclusively or alongside protons) are only found in prokaryotes, with examples in bacteria, such as *Thermotoga maritima* (Belogurov *et al.*, 2005; Pérez-Castiñeira *et al.*, 2001), and archae, such as *Methanosarcina mazei* (Malinen *et al.*, 2007). Neither plants nor protozoan parasites contain Na^+-PPases or Na^+/H^+-PPases. H^+-PPases are not exclusive and can be found in plants, protozoan parasites and prokaryotes, as can both K^+-dependent and K^+-independent variants of M-PPases. Overall, this means there are 5 types of M-PPase: K^+ -dependent H^+ -PPases, K^+ -independent H^+ -PPases, Na^+ -PPases, Na^+-dependent Na^+/H^+-PPases and true Na^+/H^+ -PPases. These differences are key to understanding how these proteins function and the respective role they have in each class of organism.

1.2.2 M-PPases in Prokaryotes

As discussed earlier, prokaryotes are the only group of organisms to contain all variants of M-PPases, and are the only group in which the Na^+-PPases and Na^+/H^+-PPases can be found. In these organisms, the M-PPase is generally located in the cell membrane but can also be found in the membranes of acidocalcisome-like organelles (Seufferheld *et al.*,

2003). The substrate always binds on the cytoplasmic side of the protein, pumping Na^+/H^+ either into the periplasmic space across the prokaryotic inner membrane or into organelles.

Thermotoga maritima is a hyperthermophilic marine bacterium that thrives near undersea volcanic vents. It has an optimal growth temperature of between 70 ˚C and 80 ˚C, with a maximum growth temperature of 90 ˚C (Huber *et al.*, 1986). The M-PPase of *T. maritima* (TmPPase) was first characterised in 2001 (Pérez-Castiñeira *et al.*, 2001) and originally thought to pump protons. Later studies (Belogurov *et al.*, 2005) found it had an absolute requirement for Na^+ and was more active in the presence of K^+. Using steady-state kinetics it was shown that increasing Na^+ concentrations led to increasing hydrolytic activity up to 10 μmol min^{-1} mg^{-1} in the absence of K^+, but had an activity of 20 μmol min^{-1} mg^{-1} in the presence of 50 mM K^+. The role of this protein in the native organism has not been studied, but given the organisms proximity to undersea thermal vents, a low energy environment, it would need to maximise the use of the available energy. This would be achieved by utilising PP_i (a waste product of ATP-hydrolysis) as a secondary source of energy. The protein itself has been used for biochemical and structural studies (Kellosalo *et al.*, 2012; Tsai *et al.*, 2014) due to its high thermal stability, which allows for high temperature solubilisation, efficient purification and increased stability during room temperature assays.

Pyrobaculum aerophilum is a species of thermophilic archaea found in extreme-temperature marine environments. As with *Thermotoga maritima*, this organism can survive extremely high temperatures, with optimal growth conditions around 100 degrees Celsius (Völkl *et al.*, 1993). It is unique amongst other species of *Pyrobaculum* in that it can also respire aerobically alongside the usual anaerobic respiration indicative of this genus (Völkl *et al.*, 1993). The M-PPase of *P. aerophilum* (PaPPase) was identified in 1999 (Drozdowicz *et al.*, 1999) where it was found to be a H^+-PPase. Moreover, it was determined to be insensitive to K^+, making it one of the first of the K^+-independent H^+-PPases to be discovered. It provided a means of comparing the different types of M-PPases, helping to determine the conserved core residues across all M-PPases. As with TmPPase, this protein has not been studied in the native organism, however it is logical to assume it fulfils a similar role to that of TmPPase, allowing the organism to survive

the extreme environment of undersea thermal vents by providing an extra means of generating proton gradients when in ATP-deficient environments.

1.2.3 M-PPases in Protozoan Parasites

In protozoan parasites, only the H^+-PPases are present, although both K^+-dependent and K^+-independent variants can occur in the same species (McIntosh *et al.*, 2001). This is characteristic of both parasites and plants, which have examples of both types of M-PPase within a single species. This is likely due to the evolutionary similarities between many parasites and certain plant species, such as red algae (Lim & McFadden, 2010; McFadden, 2000). However, the purpose of having two different variants has not been determined, possibly providing better regulation and control of the processes under different environmental and cellular conditions. In parasites, M-PPases are localised primarily in the acidocalcisome organelle (Scott *et al.*, 1998), although some are present at significantly lower levels in the Golgi apparatus and parasite cell membrane (Martinez *et al.*, 2002).

There are two broad categories of protozoan parasites that are important for studying M-PPases. The first are the Apicomplexan phylum, including species such as *Plasmodium falciparum* and *Toxoplasma gondii*, the causative agents of malaria and toxoplasmosis. These are genetically distinct from other parasites and contain a unique organelle, the apicoplast (Lim & McFadden, 2010), believed to derive from the secondary endosymbiosis of photosynthetic algae and in which a truncated set of genomic DNA is still present. Some apicomplexan species uniquely contain both K^+-dependent and K^+-independent variants of the M-PPase, most notably the *Plasmodium* species, in which the M-PPases are commonly referred to as VP1 (Vacuolar Pyrophosphatase type 1, K^+-dependent) and VP2 (Vacuolar Pyrophosphatase type 2, K^+-independent), respectively.

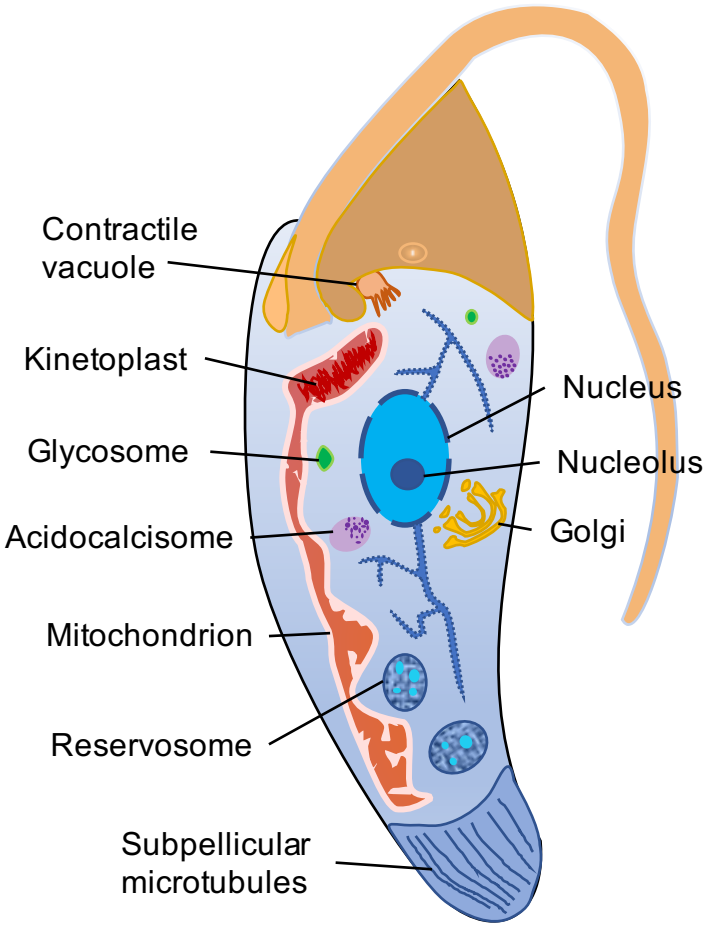


Figure 1.5: Overview of a Trypanosomatid parasite.
Schematic representation of the epimastigote form of a generalised trypanosomatid parasite showing various organelles, including the membrane pyrophosphatase-containing acidocalcisome organelle. Figure redrawn from figure 1 of (Docampo et al., 2005).

The second major group of parasites are the kinetoplastids, which include species of *Trypanosoma* and *Leishmania*, the causative agents of African sleeping sickness and Leishmaniasis, respectively. The Kinetoplastids are flagellated protozoan parasites, making them morphologically distinct from the Apicomplexans (Figure 1.5). The main distinguishing feature with this class is the presence of the Kinetoplast, a dense granule of DNA located within the single mitochondrion of these parasites that contains many copies of the mitochondrial genome (Shapiro & Englund, 1995) (Figure 1.5). The only known type of M-PPase in these parasites is the K^+-dependent H^+-PPase found in many species of *Leishmania* (Rodrigues *et al.*, 1999) and *Trypanosoma* (Scott *et al.*, 1998).

1.2.3.1 Role of the Acidocalcisome in Protozoan Parasites

The first parasites in which an acidocalcisome was discovered were *Trypanosoma* species (Docampo *et al.*, 1995) and subsequently in a range of other protozoan parasites, invariably linked to the discovery of M-PPases in these organisms (Moreno & Zhong, 1996; Pérez-Castiñeira *et al.*, 2002; Scott *et al.*, 1998). It is an acidic organelle, initially found to be an intracellular calcium store (Docampo *et al.*, 1995) vital for normal cellular functions in parasites. It was later found to be a store for a wider range of metal ions and also phosphorous in the form of phosphate, pyrophosphate and polyphosphate (Docampo & Moreno, 2011). In parasites, this organelle is also capable of regulating intracellular pH and is vital for osmoregulation due to its interaction with another organelle, the contractile vacuole (Figure 1.5). It was shown that a complex formed between these two organelles, with a resulting increase in acidocalcisome volume during periods of hypo-osmotic stress, but only once the acidocalcisome had been energised via one of several integral-membrane cation pumps (Rohloff *et al.*, 2004). Disruption of normal acidocalcisome function was shown to prevent its role in maintaining cell homeostasis during periods of environmental stress (Lemercier *et al.*, 2002; Lemercier *et al.*, 2004). This is important due to extreme changes in the osmolarity of the external environment as the parasite transitions from the insect mid-gut, at low pH and high osmolarity to the host organism, which is a neutral pH and very low osmolarity.

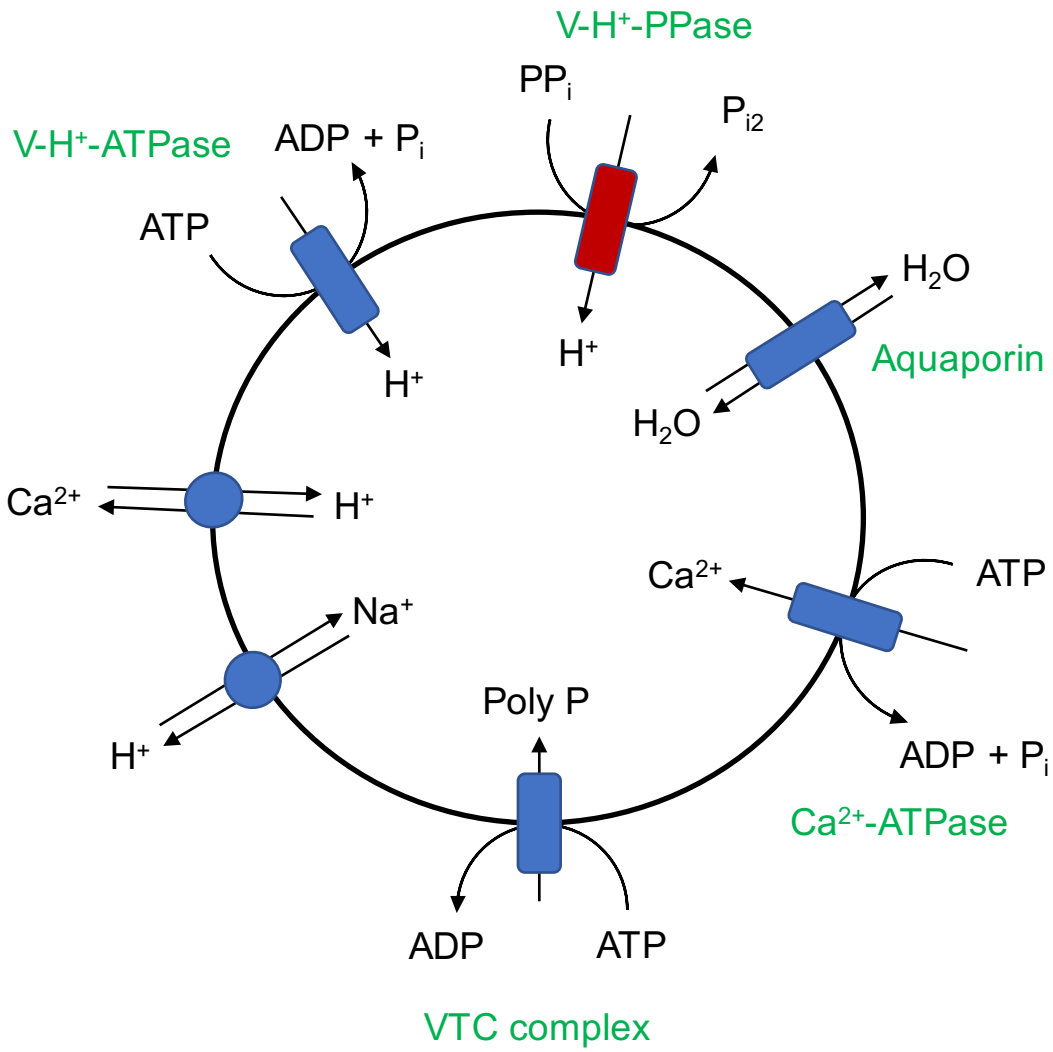


Figure 1.6: Representation of Acidocalisome Transporters and Channels.
Schematic overview of some of the transporters, channels and pumps of the acidocalcisome organelle found in protozoan parasites. Only key proteins are shown, including the H^+-ATPase and H^+-PPase (highlighted in red). Figure redrawn and simplified from Figure 4 of (Docampo et al., 2005)

The acidocalcisome contains a range of transporters, including the H^+-PPase and a V-type H^+-ATPase, which also appears to energise the acidocalcisome via proton translocation (Docampo & Moreno, 2011) (Figure 1.6). Despite this, several studies have shown the importance of these proteins to parasite survival and infectivity. One study used targeted RNA interference in *Trypanosoma bruceii* to generate inducible knockouts of the H^+-PPase (Lemercier *et al.*, 2002) in procyclic and blood-stage forms of the parasite. The acidocalcisomes lost acidity, 90 % of the polyphosphate and could not release Ca^{2+}. Moreover, the parasites could not adapt to changes in the extracellular pH, were slower to recover their intracellular pH and the parasites grew at a half the rate and only reached a cell density one fifth that of the wild type parasites.

Knocking out the M-PPase of *Toxoplasma gondii* left the tachyzoite stage of the parasites unable to attach to or infect new cells as well as disrupting osmoregulation (Liu *et al.*, 2014). This loss of infectivity implies that M-PPases play an important role in driving infectivity in these parasites via their activation of the acidocalcisome, alongside the various ATPases and highlights the importance of these proteins as potential drug targets.

1.2.4 M-PPases in Plants

Plants contain similar variants of M-PPases as those found in protozoan parasites, including both the K^+-dependent and K^+-independent variants of the H^+-PPases. These are located in the vacuole and are particularly important for the survival of cells during abiotic stresses, such as during periods anoxia, coldness (Carystinos *et al.*, 1995), high soil salinity (Gao *et al.*, 2006) or drought (Brini *et al.*, 2007; Venancio *et al.*, 2014). In both these cases, overexpression of the M-PPase was found to greatly enhance the survival and viability of plants, leading to better growth compared to plants that had low or normal levels of the protein (Pasapula *et al.*, 2011).

Alongside the established stress-response role, the H^+-PPase has also been strongly linked to several developmental roles in plants. Auxin transport is a major factor in plant organ development and general plant growth (J. Li *et al.*, 2005). The overexpression of H^+-PPase was linked to the increase in auxin transport mechanisms and subsequently an increase in plant organ development and general growth. Another study of *Arabidopsis* development showed that if the H^+-PPase was knocked out, the plants failed to grow post-germination due to a toxic build-up of cytosolic PP_i (Ferjani *et al.*, 2011). It was possible to rescue the phenotype via expression of the S-PPase, suggesting a much greater and more general role for PPases in plant development.

Vigna radiata, more commonly known as the mung bean, is another plant that has been studied in great detail with regards to M-PPases. It is a member of the legume family of plants and an important food crop in Southern and Eastern parts of the world and has been shown to contain a K^+-dependent H^+-PPase (Shen-Hsing *et al.*, 2009). The structure of this protein was determined via X-ray crystallography (Lin *et al.*, 2012), making it one of the first structures of an M-PPase to be solved, alongside the Na^+-PPase from *Thermotoga maritima* (Kellosalo *et al.*, 2012).

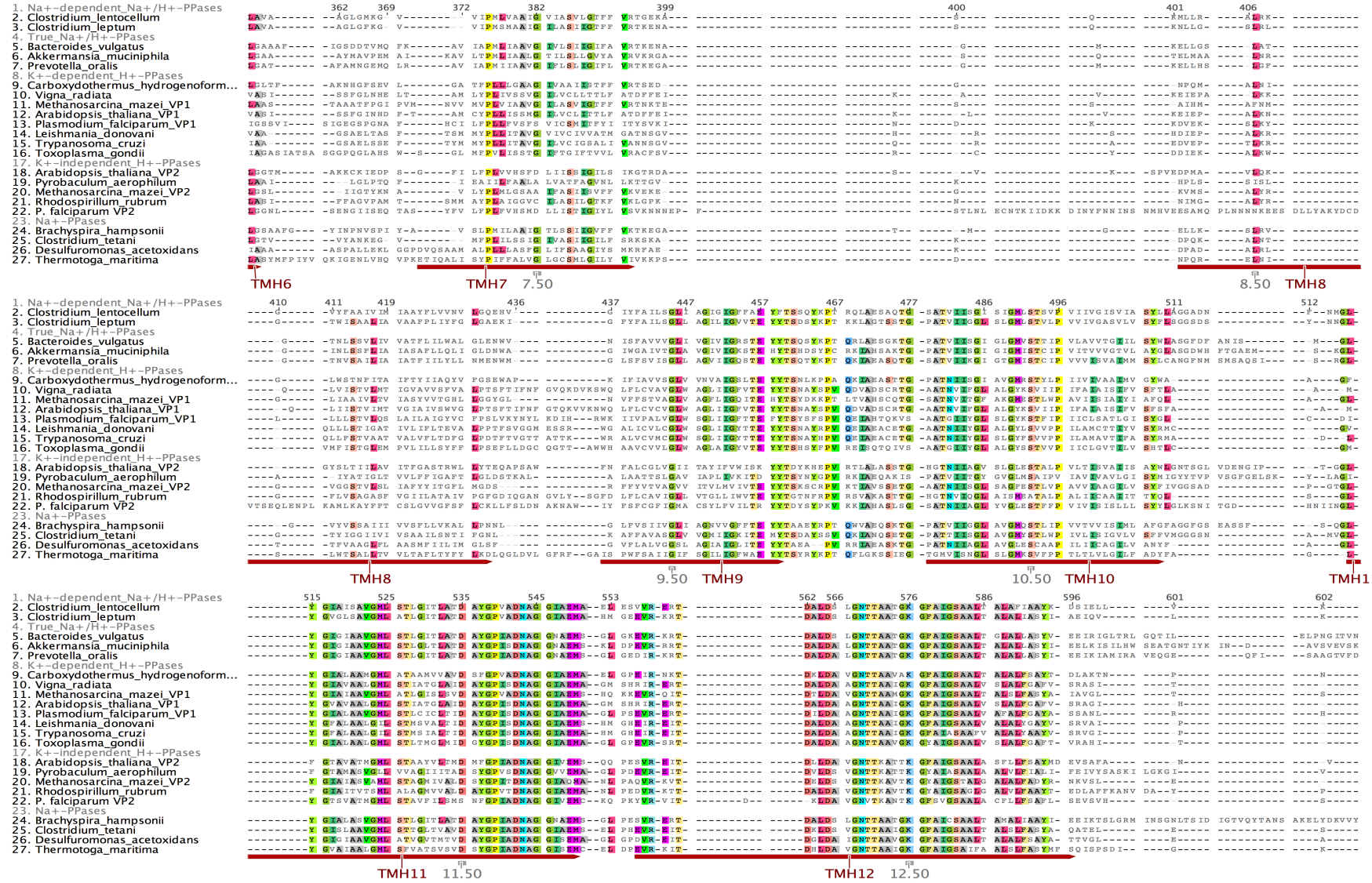
1.2.5 Evolution of M-PPases

The evolutionary lineage of M-PPases has been examined in detail to try to better group these proteins based on function and phylogenetic origin. It was proposed that the Na^+-PPases were the first type of M-PPase, originating from early Na^+ transporters, with the transition from Na^+ to H^+ occurring over time (Luoto *et al.*, 2011). Some diverged into the dual Na^+/H^+-pumping and others became the H^+-pumping PPases. The prevailing hypothesis is that each variant diverged separately from the early Na^+-PPases rather than emerging gradually over time along a single evolutionary path. This hypothesis is based on the evolutionary origins of the various types of ATPases (Mulkidjanian *et al.*, 2008), whereby sodium-pumping preceded proton-pumping, due to difficulties for early proteins to maintain proton-impermeable membranes compared with sodium-impermeable ones.

An alignment of several sequences from each type of M-PPase (Figure 1.7) revealed a high degree of similarity between the M-PPases of different species. The highest similarity was seen for the 6 inner-ring helices (5, 6, 11, 12, 15 and 16) and is independent of the type of M-PPase. The difference between K^+-dependent and K^+-independent variants is related to residues at position 12.46 and 12.49, which are invariably Lysine and Threonine in K^+-independent variants, but are Alanine and Glycine/Alanine in K^+-dependent variants (Figure 1.7). Even using the sequence alignment, it has not been possible to ascertain the specific residues involved with cation specificity in M-PPases.

The parasite forms of M-PPases are more closely related to the plant M-PPases than any other variant, given that the sequence similarity is around 50-60%, compared with around a 30% similarity to bacterial forms of the protein. This link between plants and parasites has been seen with numerous other proteins, including those found in the apicoplast organelle of apicomplexan parasites due to the evolutionary link between protozoan parasites and species of algae (McFadden, 2000). This relationship between plants and protozoan parasites may also explain why these are the only eukaryotic organisms to possess M-PPases. The role of M-PPases in plant development and stress responses (discussed above) shows why they are essential for these organisms, with presence of M-PPases in parasites resulting from the common evolutionary lineage of these species.


1. Na+-dependent_Na+/H+-PPases
2. Clostridium_lentocellum
3. Clostridium_leptum
4. True_Na+/H+-PPases
5. Bacteroides_vulgatus
6. Akkermansia_muciniphila
7. Prevotella_oralis
8. K+-dependent_H+-PPases
9. Carboxydothermus_hydrogenoform...
10. Vigna_radiata
11. Methanosarcina_mazei_VP1
12. Arabidopsis_thaliana_VP1
13. Plasmodium_falciparum_VP1
14. Leishmania_donovani
15. Trypanosoma_cruzi
16. Toxoplasma_gondii
17. K+-independent_H+-PPases
18. Arabidopsis_thaliana_VP2
19. Pyrobaculum_aerophilum
20. Methanosarcina_mazei_VP2
21. Rhodospirillum_rubrum
22. P. falciparum VP2
23. Na+-PPases
24. Brachyspira_hampsonii
25. Clostridium_tetani
26. Desulfuromonas_acetoxidans
27. Thermotoga_maritima
TMH1
1.50
2.50
TMH2
3.50
TMH3
TMH4
4.50
TMH4
5.50
TMH5
5-6 border
TMH6
6.60


1. Na+-dependent_Na+/H+-PPases
2. Clostridium_lentocellum
3. Clostridium_leptum
4. True_Na+/H+-PPases
5. Bacteroides_vulgatus
6. Akkermansia_muciniphila
7. Prevotella_oralis
8. K+-dependent_H+-PPases
9. Carboxydothermus_hydrogenoform...
10. Vigna_radiata
11. Methanosarcina_mazei_VP1
12. Arabidopsis_thaliana_VP1
13. Plasmodium_falciparum_VP1
14. Leishmania_donovani
15. Trypanosoma_cruzi
16. Toxoplasma_gondii
17. K+-independent_H+-PPases
18. Arabidopsis_thaliana_VP2
19. Pyrobaculum_aerophilum
20. Methanosarcina_mazei_VP2
21. Rhodospirillum_rubrum
22. P. falciparum VP2
23. Na+-PPases
24. Brachyspira_hampsonii
25. Clostridium_tetani
26. Desulfuromonas_acetoxidans
27. Thermotoga_maritima
TMH6
TMH7
7.50
8.50
TMH8
9.50
TMH9
10.50
TMH10
TMH11
11.50
TMH12
12.50

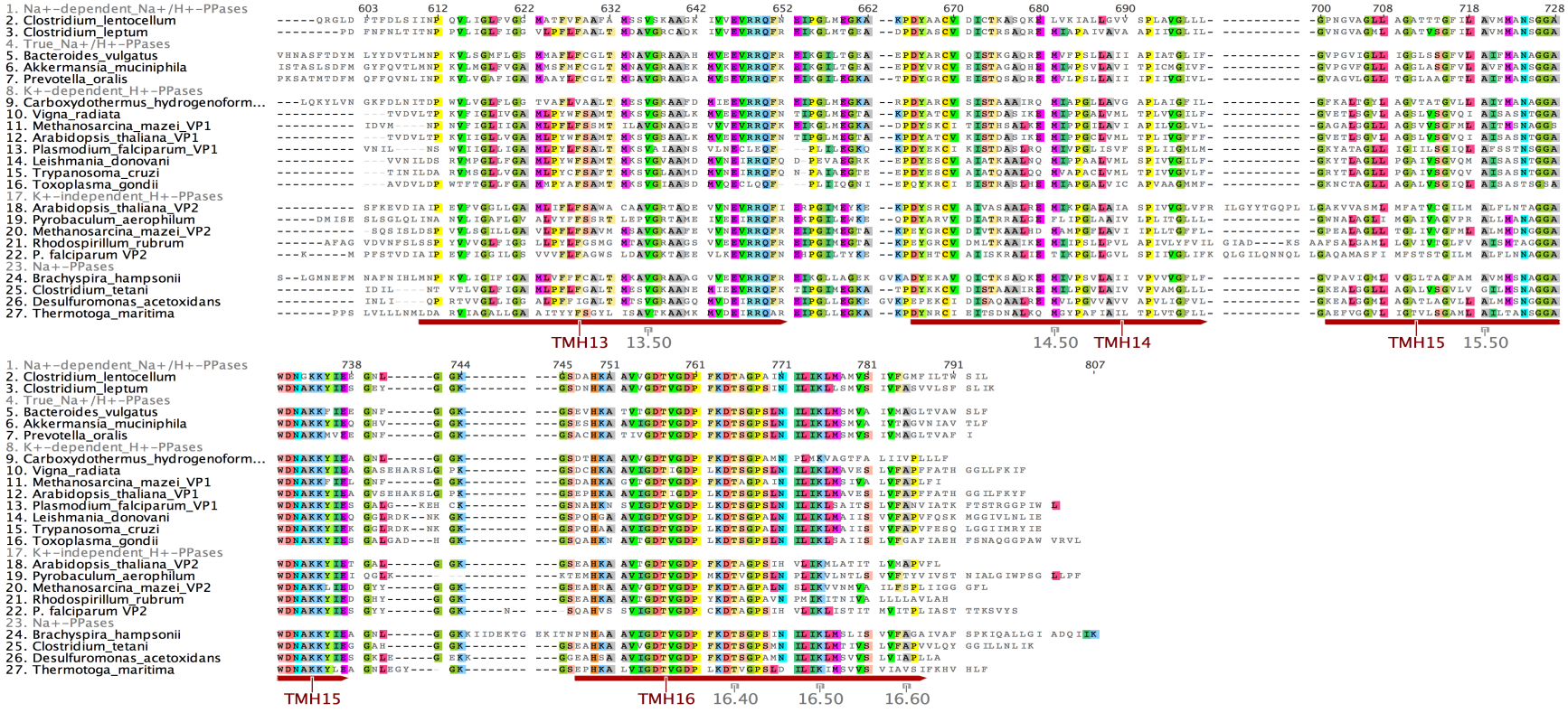


Figure 1.7: Sequence Alignment of Membrane Pyrophosphatases.

Full sequence alignment for several examples of each main type of M-PPase (Na^+-PPase, K^+-dependent H^+-PPase, K^+-independent H^+-PPase, Na^+-dependent Na^+/H^+-PPase and true Na^+/H^+-PPase). Transmembrane regions are labelled based on TmPPase structures (PDB ID: 4AV3 and 4AV6) along with the conserved anchor residues used for Ballesteros and Weinstein numbering of M-PPases (See Section 1.4).

1.3 M-PPases as Drug Targets

The M-PPases of protozoan parasites are essential to infectivity (see above) and are thus a potential new drug target for combating deadly parasitic infections such as malaria, toxoplasmosis, leishmaniasis and trypanosomiasis (Docampo & Moreno, 2008). The treatments for these diseases can present serious challenges, with a sharp rise in resistance to many drugs currently used. This is seen for the treatment of trypanosomiasis, where there is increasing resistance for the few drugs available (Delespaux & de Koning, 2007). Treatments for malaria are also hampered in a similar way, with drug resistance to frontline malarial drugs, such as chloroquine, pyrimethamine, quinine, mefloquine and artemisinin, reaching dangerously high levels (Dondorp *et al.*, 2009; Fairhurst & Dondorp, 2016; Yuvaniyama *et al.*, 2003).

Drug design is often hindered by the intracellular localisation of protozoan parasites in host cells during infection, as it is difficult to get drugs across multiple membranes and remain active. These parasites also contain highly efficient efflux mechanisms that rapidly remove active drugs from the cytoplasm or cellular compartments. An example of this is the main mechanism of chloroquine resistance in *Plasmodium* parasites, where a mutation in the acidocalcisome transporters increases the active secretion of chloroquine before it can reach toxic levels (Krogstad *et al.*, 1987). M-PPases represent a potential drug target to combat the rise in drug resistance towards malarial parasites. Drugs designed against these targets should significantly reduce the infectivity and longevity of the parasites in host cells. Targeting of the acidocalcisome via the M-PPases would also diminish the parasites ability to cope with the sudden change in pH or osmolarity and interfere with calcium-mediated signalling pathways, described above (Section 1.2.3.1).

Studies on PP_i pathways in various organisms have highlighted their importance and the potential for new drug targets. A previous study showed the viability of drug development against the PP_i regulatory pathways of protozoan parasites (Francia *et al.*, 2011). This study looked at the NHE3 protein of *Toxoplasma gondii*, which was found to co-localise with the H^+-PPase. When TgNHE3 was knocked out, the levels of PP_i were greatly reduced, preventing the H^+-PPase from functioning efficiently. This left the cells significantly more susceptible to hypo-osmotic shock and led to toxic levels of

extracellular sodium and calcium. Although not looking at the effects of specifically knocking out the H^+-PPase, it highlights the importance of PP_i in cells and shows that disruption of the related pathways can be fatal to the parasites during transition from the mid-gut of the insect vector to the host bloodstream, due to a rapid increase in pH and decrease in osmolarity.

Several studies have also looked specifically at using M-PPases as drug targets, including the H^+-PPase from the parasitic species of ringworm, *Baylisascaris schroederi*. This has been evaluated for use as part of a potential vaccine. Other studies on H^+-PPases in *T. gondii* tachyzoites (Rodrigues *et al.*, 2000) showed that by inhibiting PP_i hydrolysis and proton pumping led to an increase in the concentration of intracellular PP_i. The level increased significantly above that of ATP and inhibited the growth of tachyzoites in tissue culture cells. Experiments conducted using inhibitors of PP_i hydrolysis and proton pumping showed that the tachyzoites could not grow in cell cultures. Based on these findings, it is clear that M-PPases are viable drug targets for protozoan parasites in various stages of the life-cycle.

1.4 Structure and Function of M-PPases

1.4.1 Ballesteros and Weinstein Numbering System

During structural studies, it is useful to find a means of conveying uniformity in residue numbering, which often varies greatly between different species. The Ballesteros and Weinstein numbering system has been adopted for use with the M-PPases to try to solve this problem. This system has been extensively used to compare G-protein coupled receptors (Ballesteros & Weinstein, 1995) and allows for easy comparison of residues. This system labels each residue thusly: $X^{a.b}$, where 'X' represents the one letter code for the amino acid, 'a' represents the helical number on which the residue is present and 'b' represents the position upon that helix based on a highly conserved anchor residue common to all known M-PPases. In this way, the residues can be compared with much greater convenience.

1.4.2 Previous Structural Studies of M-PPases

So far, only three structures have been solved of M-PPases from two different species. These are the structure of a H^+-PPase in plant cells from *Vigna radiata* (VrPPase) (Lin *et al.*, 2012) and the structure of an Na^+-PPase from *Thermotoga maritima* (TmPPase) (Kellosalo *et al.*, 2012).

1.4.2.1 Expression, Solubilisation and Purification of M-PPases

The expression and purification of M-PPases was difficult, as it is with many integral-membrane proteins. Poor yields, difficulty maintaining stability when solubilising the protein into detergent micelles and difficulties in obtaining a pure protein product are common issues when working with membrane proteins. Despite this, expression was achieved for the M-PPases of *Arabidopsis thaliana* (Kim *et al.*, 1994), *Thermotoga maritima* (Pérez-Castiñeira *et al.*, 2001) and *Vigna radiata* (Hsu *et al.*, 2009), all of which were expressed in yeast.

In 2005, a new solubilisation and purification technique was utilised for M-PPases taking advantage of the thermostable properties of the M-PPase from *T. maritima*. This method was referred to as the "hot-solve" purification technique (López-Marqués *et al.*, 2005) and involved solubilising the expressed protein in detergents at an extremely high temperature (70-80 ˚C). This denatured any contaminating proteins, leaving only the thermostable TmPPase to bind to the Ni-NTA column. This resulted in high yields of pure protein samples suitable for testing and essential for crystallographic studies.

1.4.2.2 Crystallisation of Membrane Proteins

Crystallographic studies can be carried out using relatively high concentrations of protein, typically around 10 mg/ml and so requires high expression yields to carry out vast crystal screens. Early stages involve crystallisation trials, where various buffer and environmental conditions are tested to find ones that lead to crystal formation. These conditions generally involve the type and concentration of buffer and various salts or precipitants alongside changes in temperature and pH.

Membrane proteins are unique in that they must be solubilised in a detergent micelle, prior to structural studies, which mimics the native membrane environment, proving

hydrophobic contact with the protein and hydrophilic contact with the surrounding solvent (Garavito & Ferguson-Miller, 2001). Micelles ensure the protein remains folded and stable in solution, but hinders crystallisation by reducing the number of crystal contacts. These contacts are required to form stable and well-ordered crystals and so detergent screens are carried out during crystallisation trials.

The critical micellar concentration (CMC) determines the absolute minimum concentration of a detergent that is required to form stable micelles. Varying the concentration of detergent above the CMC can have significant impacts on the crystallisation. In addition, testing different types of detergent can also have significant effects on the crystallisation, as the chain length of various detergents can affect the size of the resulting micelle (Oliver *et al.*, 2013). Smaller detergent micelles allow for more crystal contacts to form, but the smaller-chained detergents tend to be more denaturing to the protein (Garavito *et al.*, 1996). This means that many detergents must be screened to find the optimal one that will maintain the stability of the protein with as small a micelle size as possible.

Membrane protein crystallisation can also generate detergent crystals, alongside salt crystals found in general protein crystallography. As with salt crystals, detergent crystals can be detected either using UV-imaging of the crystals, which reveals a negative fluorescence (detergent crystals appear darker than the well solution), or by the distinctive single-ring X-ray diffraction pattern. Detergent crystals also have a fairly distinct morphology, appearing more rounded than salt or protein crystals.

Data can be collected by subjecting the crystal to an X-ray or synchrotron beam and measuring the resulting diffraction pattern, consisting of a series of clearly defined spots. Crystals are rotated within the beam in small increments and multiple diffraction images are generated. In order to obtain the structure factors from the diffraction images, two pieces of information are required, the amplitude and phases. The amplitudes can be obtained from the intensities of the diffraction spots during indexing, allowing for calculation of the space group and unit cell parameters. The phases can be determined either experimentally or using molecular replacement.

Experimental phase calculation can be carried out using heavy atom soaks, which alter the intensities of the diffraction spots. Comparisons between the normal dataset and the heavy-atom dataset allow for the determination of the heavy atoms positions within the protein, providing the necessary phase information. The alternative method is molecular replacement, in which the phases of a structure can be obtained from a related structure that has been previously solved. If such a structure exists, then the phases can be used to acquire the structure factors for the new dataset. The resulting structure factors can be used to generate an electron density map, in which the model of the protein can be built. This is refined against the original data to ensure that that model best fits the available data.

1.4.2.3 M-PPase Crystallisation and Structure Solution

The crystal structures of both TmPPase and VrPPase have been solved in a range of catalytic states. TmPPase has been solved in the resting state with magnesium and calcium ions occupying the hydrolytic centre (PDB ID: 4AV3)(Kellosalo *et al.*, 2012) and also in a product-bound state with two phosphates and four magnesium ions in the active site (PDB ID: 4AV6)(Kellosalo *et al.*, 2012). In addition, the structure of VrPPase was solved in complex with Imidodiphosphate (IDP) (PDB ID: 4A01)(Lin *et al.*, 2012), an analogue of PP_i that cannot be hydrolysed by the enzyme. This inhibits the protein by locking it into the substrate-bound form, allowing structures of this state to be observed. These structures have formed the basis for our current understanding of the structure, function and catalytic cycle of M-PPases.

1.4.2.4 Overview of M-PPase Structures

The overall structure of M-PPases appears to be common for both TmPPase and VrPPase and thus both H^+ and Na^+-PPases. The protein forms into a homodimer, with each protomer consisting of 16 transmembrane helices split into 2 observable ring structures (Figure 1.8a) (Kellosalo *et al.*, 2012; Lin *et al.*, 2012). The outer ring structure (helices 1-4, 7-10 and 13-14) forms the main dimer interface and interacts with the membrane, housing an inner ring of 6 helices (helices 5-6, 11-12 and 13-14) forming the central channel. This channel contains four distinct regions: the hydrolytic centre, a coupling channel connecting this to the ion gate and the exit channel (Figure 1.8b). The hydrolytic centre is located on the cytoplasmic side of the protein, where the available pyrophosphate

is found. It is in this region that the hydrolysis of PP_i into P_{i2} occurs. Below this is the coupling channel, containing a series of charged residues, where it is proposed that ion/proton translocation occurs.

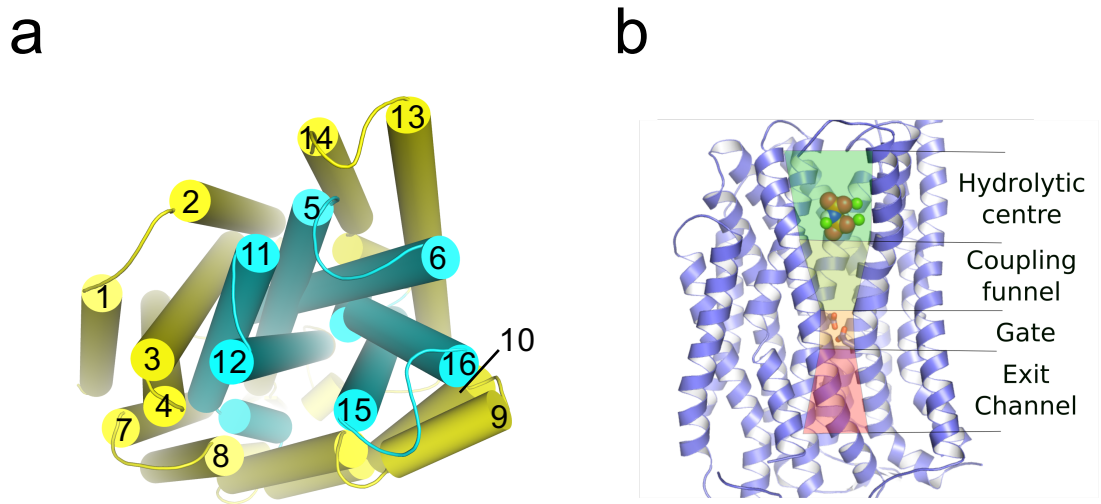


Figure 1.8: Structural Overview of Membrane Pyrophosphatase Channel.
M-PPase monomer (a) viewed from the cytoplasmic side showing the two-ring composition, with inner ring (helices 5, 6, 11, 12, 15 and 16) forming the central channel and (b) viewed from the side, showing the four-part composition of the channel, with the hydrolytic centre at the cytoplasmic side through to the exit channel on the exterior side of the membrane.

The ion gate (Figure 1.8) ensures directionality of pumping by preventing backflow of ions. All three available structures show the gate in the closed conformation, as it only opens in time-scales too rapid to be detected via crystallography. Below the gate is the exit channel that, once opened, will release the pumped ion to the outside. Following hydrolysis, the two phosphate molecules are sequentially released back into the cytoplasmic side of the membrane, as shown by the TmPPase:Mg_4P_{i2} and VrPPase:Mg_2P_i structures.

1.4.3 M-PPase Mechanism of Action

The structure of the VrPPase protein in the substrate-bound form reveals the position of the substrate-analogue (IDP) alongside the various metal ions that coordinate the substrate around the nearby residues (Lin *et al.*, 2012). Several key residues were identified, including a series of aspartate, lysine and arginine residues (Figure 1.9). The resolution was also sufficient to observe a bound potassium ion in the active site alongside numerous water molecules, including one that was proposed to act as the attacking nucleophile. Unlike the type-I and II S-PPases, this potential nucleophile was not found

to be coordinated by the metal ions, but instead by two aspartate residues $D^{6.43}$ and $D^{16.39}$ located directly beneath the active site (Lin *et al.*, 2012), more akin to the type-III active site coordination. In M-PPases, the metal ions play a role in the catalysis, by coordinating the substrate, stabilising the leaving group phosphate and negating the charge on the electrophilic phosphate group, allowing hydrolysis to take place via the nucleophilic water (Heikinheimo *et al.*, 1996). The specific mechanism behind the activation of this water nucleophile has not been studied in M-PPases so currently not much information is known.

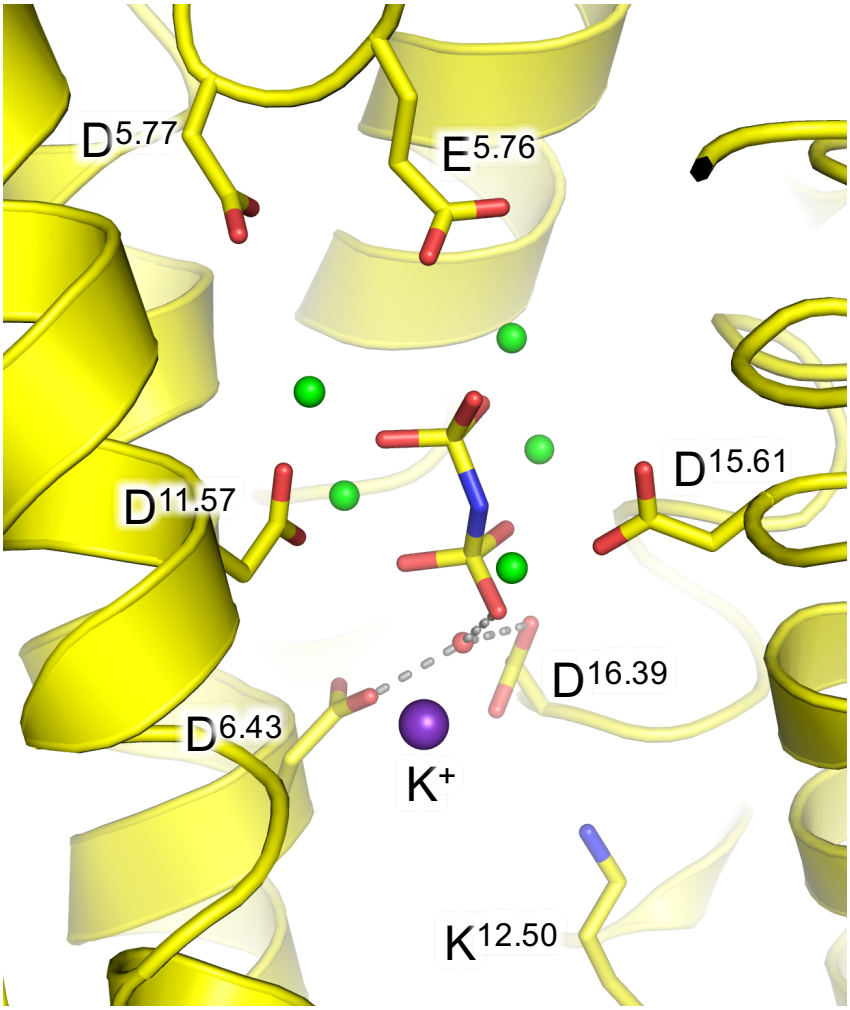


Figure 1.9: Overview of the Membrane Pyrophosphatase Hydrolytic Centre.
Overview of the hydrolytic centre of the IDP-bound VrPPase structure (VrPPase:Mg_5IDP, PDB ID: 4A01) showing bound Mg_5IDP complex, bound K^+, loop 5-6 packed over the active site and coordination of the nucleophilic water (red sphere) by $D^{6.43}$ and $D^{16.39}$.

1.4.3.1 Role of Metal Ions

Various combinations of metal ions have been found in the active site of each of the crystal structures. Given the functional similarities between the mechanism of hydrolysis in M-PPases to those found in S-PPases, it was proposed that there would be a similar requirement for metal ions in the activity of M-PPases. It was determined that a series of magnesium or manganese ions formed a complex with PP_i, similar to that seen in the S-PPases (Cooperman *et al.*, 1981; Heikinheimo *et al.*, 1996). Two of these ions bound to the enzyme itself, as seen in the resting state structure of TmPPase (one metal ion being Ca^{2+} in the crystal structure). Mg^{2+} was found to be the most potent activator of PPase activity but manganese could also activate the protein (Maeshima, 1991). The same study found that calcium was a potent inhibitor of M-PPase and once bound, would prevent the $MgPP_i$ substrate from binding. In this way, it is possible to lock the protein into the resting-state conformation with calcium. The role of Mg^{2+} was described in section 1.4.3. In addition to the metal ions in the active site, there are additional ions that affect the activity of various types of M-PPase, such as K^+ and Na^+.

1.4.3.2 Potassium Ion Dependence

Two sodium binding sites were discovered in TmPPase (Belogurov & Lahti, 2002), one of which exclusively binds sodium. The other site can also bind potassium and in doing so increases the affinity for sodium ions at the other site. Mutants of the the ion gate were generated and it was determined that $D^{16.46}$ was strongly involved in the binding of sodium to TmPPase and likely forms part of the binding site in this protein (Belogurov & Lahti, 2002). No additional information on the sodium-exclusive binding site was obtained from the TmPPase structures, however, the Na^+/K^+ site was identified in the VrPPase structure.

K^+ is visible in the active site at the proposed position near $A^{12.46}$, providing a potential mechanistic explanation for K^+-dependence. In K^+-independent M-PPases this residue is a lysine. It is therefore proposed that the side chain amine group of the lysine occupies the same position as the potassium ion, substituting the positive charge. Indeed, this hypothesis is backed up by mutational studies in *Carboxydothermus hydrogenoformans*, a K^+-independent H^+-PPase (Belogurov & Lahti, 2002), in which $A^{12.46}K$ and $A^{12.49}T$ mutants were generated. The $A^{12.46}K$ mutant was sufficient to abolish the K^+-sensitivity,

with the resulting protein equally active in the presence of increasing K^+ concentrations from 0 to 50 mM. The $A^{12.49}T$ mutation did not generate a K^+-independent protein, but instead increased the affinity for K^+ three-fold.

1.4.3.3 Cation Selectivity

The basis for ion selectivity (H^+ or Na^+) has not been established in M-PPases, although it is clearly complex given the high degree of sequence similarity, notably in the percentage of conserved residues around the hydrolytic centre and at the ion gate (Figure 1.7). The gate itself is comprised of either two or three charged residues. In VrPPase, $D^{6.50}$ and $K^{16.50}$ form an ion pair with a glutamate at position $E^{6.57}$ (Lin *et al.*, 2012) (Figure 1.10). In TmPPase an ion triplet is formed with $D^{6.50}$, $D^{16.50}$, $E^{6.53}$ interacting with $K^{12.50}$ (Figure 1.10). The glutamate is positioned a single turn up the helix (towards the cytoplasmic side) in TmPPase compared with VrPPase (Kellosalo *et al.*, 2012). This change in position alters the interactions that form at the ion gate and thus seems a plausible mechanism to explain the ion selectivity.

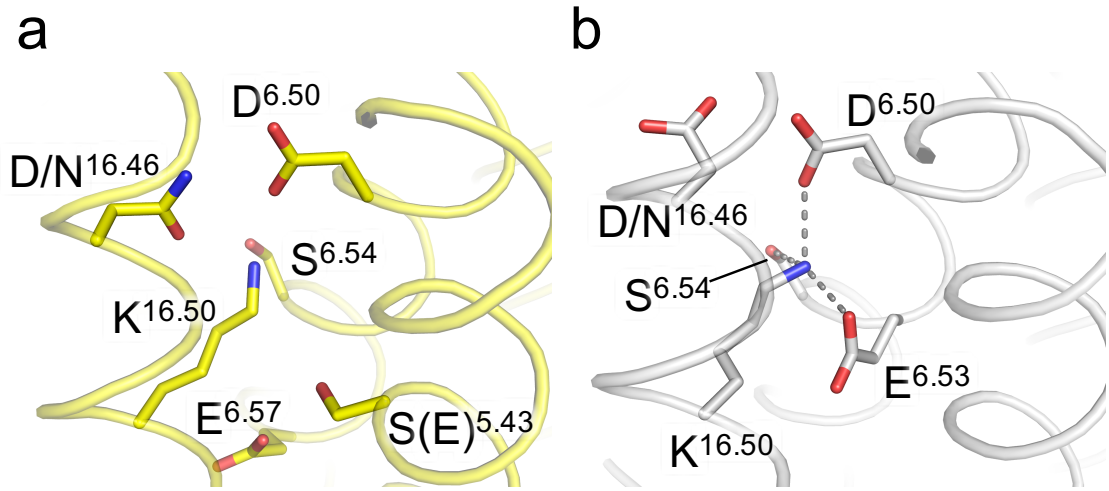


Figure 1.10: Overview of the Membrane Pyrophosphatase Ion Gate.
Overview of the ion gate of (a) the VrPPase:Mg_5IDP (yellow) and (b) TmPPase:CaMg (grey) structures showing key residues and structure of ion gate through interactions between $K^{16.50}$ and surrounding residues. Labelling of $E^{6.53}$ is based on TmPPase structure, $E^{6.57}$ is based on VrPPase structure and $S^{5.43}$ from VrPPase represents the position of $E^{5.43}$ in Flavobacterium johnsaniae M-PPase.

However, it is clear that this motif is not conserved across all types of M-PPases. Indeed, the glutamate is found at position $E^{6.53}$ for the majority of M-PPase types, both Na^+-PPases, H^+-PPase and dual-PPases, only really differing for the K^+-dependent H^+-PPases

at position $E^{6.57}$. There are some notable exceptions, such as in *Flavobacterium johnsaniae*, in which this glutamate is positioned at $E^{5.43}$. However, this is still predicted to occupy the same region in the structure based on models and mutational studies of this semi-conserved glutamate (Figure 1.10) (Luoto *et al.*, 2011). These found that if the glutamate of an Na^+-PPase is moved from $E^{6.53}$ to $E^{6.57}$, as seen in the K^+-dependent H^+-PPases, then there is a significant drop in hydrolytic activity and ion pumping, but no change in the ion selectivity is observed. Ultimately, this means there is no consensus on which residues are essential for ion selectivity.

1.5 Studies of M-PPase Catalytic Cycle

1.5.1 Catalytic Cycle from Multi-state Model

A model of the M-PPase catalytic cycle had been proposed based upon the solved structures of TmPPase and VrPPase (Kellosalo *et al.*, 2012; Lin *et al.*, 2012; Tsai *et al.*, 2014), which covered many of the important catalytic states (resting state, substrate-analogue-bound and product-bound). This highlighted several interesting features about specific helices and residues, most notably a movement of helix 12 upon substrate binding. This has adopted a ‘’downwards’’ conformation having moved by around 2 Å and appears to have disrupted an ion pair between $K^{12.50}$ and $D^{6.43}$. This aspartate residue is involved with coordinating and activating the nucleophilic water in the active site, involved with hydrolysis of the substrate (Figure 1.9).

Upon substrate binding, the 5-6 loop region closes the active site on the cytoplasmic side preventing the backflow of ions (Tsai *et al.*, 2014). The hydrolysis of substrate occurs via the activation of the nucleophilic water. Two aspartate residues ($D^{6.43}$ and $D^{16.39}$), located just below the active site, coordinate the water nucleophile (Figure 1.10). The active site is comprised of a series of charged aspartate residues interacting with the Magnesium ion cage surrounding the pyrophosphate. This leads to a complex network of electrostatic interactions, whereby there is total coordination of all lone pairs in the substrate, requiring an additional magnesium ion with the substrate compared with the type-I S-PPases. The exact trigger for hydrolysis has yet to be determined, either occurring immediately upon substrate binding or only after cation pumping.

The movement of helix 12 is also involved with the coupling of the hydrolytic mechanism to the ion pumping mechanism, based upon the structures (Tsai *et al.*, 2014) and on an extensive series of mutations made to the M-PPase of *Streptomyces coelicolor* (Hirono & Maeshima, 2009; Hirono *et al.*, 2007). By mutating residues of helix 12, it was possible to partially decouple the enzyme, reducing the pumping activity to 25 % that of wild-type whilst maintaining hydrolytic activity between 75 and 150 %. An additional study on VrPPase (Asaoka *et al.*, 2014) identified several residues of the coupling channel: $I^{12.54}$, $L^{12.64}$, $N^{16.46}$, $V^{16.54}$ and $L^{16.57}$ that, when mutated, completely prevented proton-pumping activity, but maintained hydrolytic activity between 39 and 75 % that of wild-type.

1.5.2 Chronological Ordering of Mechanism Coupling

The model of the M-PPase catalytic cycle has raised several important questions with regards to the coupling of PP_i hydrolysis with the pumping of Na^+/H^+. The most important of these is the order in which the two key mechanisms occur: substrate hydrolysis and Na^+/H^+ pumping. There are currently three competing hypotheses as to which order the two events occur in (Figure 1.11). The first was derived from the VrPPase and suggests that hydrolysis is the first step, leading to pumping of the cation (Lin *et al.*, 2012). The second hypothesis (the binding-change method) suggests that the pumping of the cation occurs first and that it is structural changes resulting from this that allows hydrolysis to occur (Kellosalo *et al.*, 2012). The third involves the direct pumping of the proton abstracted from PPi during hydrolysis (Baykov *et al.*, 2013).

In the first model (Lin *et al.*, 2012), hydrolysis of the PP_i substrate would occur first, via the nucleophilic water. The resulting ion would be transferred down the coupling channel to the ion gate. The arrival of the cation at the ion gate would trigger the opening of the gate and result in release of the cation to the external side of the membrane (Figure 1.11a). In the binding change model, the binding of Mg_2PP_i generates a transient conformational state in which both the hydrophobic gate and exit channel are open, mediated by the downward motion of helices 11-12 (Kellosalo *et al.*, 2012). This allows an ion to exit the channel, reducing the charge in the region and generating a charge difference between the ion gate and active site. This charge difference leads to an increase in the basicity of the $D^{6.43}$ and $D^{16.39}$ that in turn activates the nucleophilic water and results in the

hydrolysis of PP_i. In this method, only once the ion has been pumped will hydrolysis be possible, thus ensuring that the entire cycle is controlled and ordered (Figure 1.11b).

The third model utilised a direct-pumping method (Baykov *et al.*, 2013). The substrate would bind and be hydrolysed via the nucleophilic water, with the resulting abstracted proton being directly pumped out of the channel without passing along a Grotthuss chain mechanism (Figure 1.11c). This model is unlikely to be correct since it does not account for Na^+-PPases. The mechanism is likely to be similar for both Na^+ and H^+-PPases, given the high sequence and structural similarities at the hydrolytic centre and ion gate (Figure 1.7). The H^+-PPases would likely utilise a Grotthuss chain mechanism (Agmon, 1995) to transfer the protons down the coupling channel. This involves the movement of protons along a 'proton wire' consisting of charged residues in the coupling channel.

It is hoped that as new structures become available at a higher resolution, the exact mechanism by which the ions are transported can be fully elucidated, using a combination of structural and biochemical information to try to pin down the specific residues involved with the process.

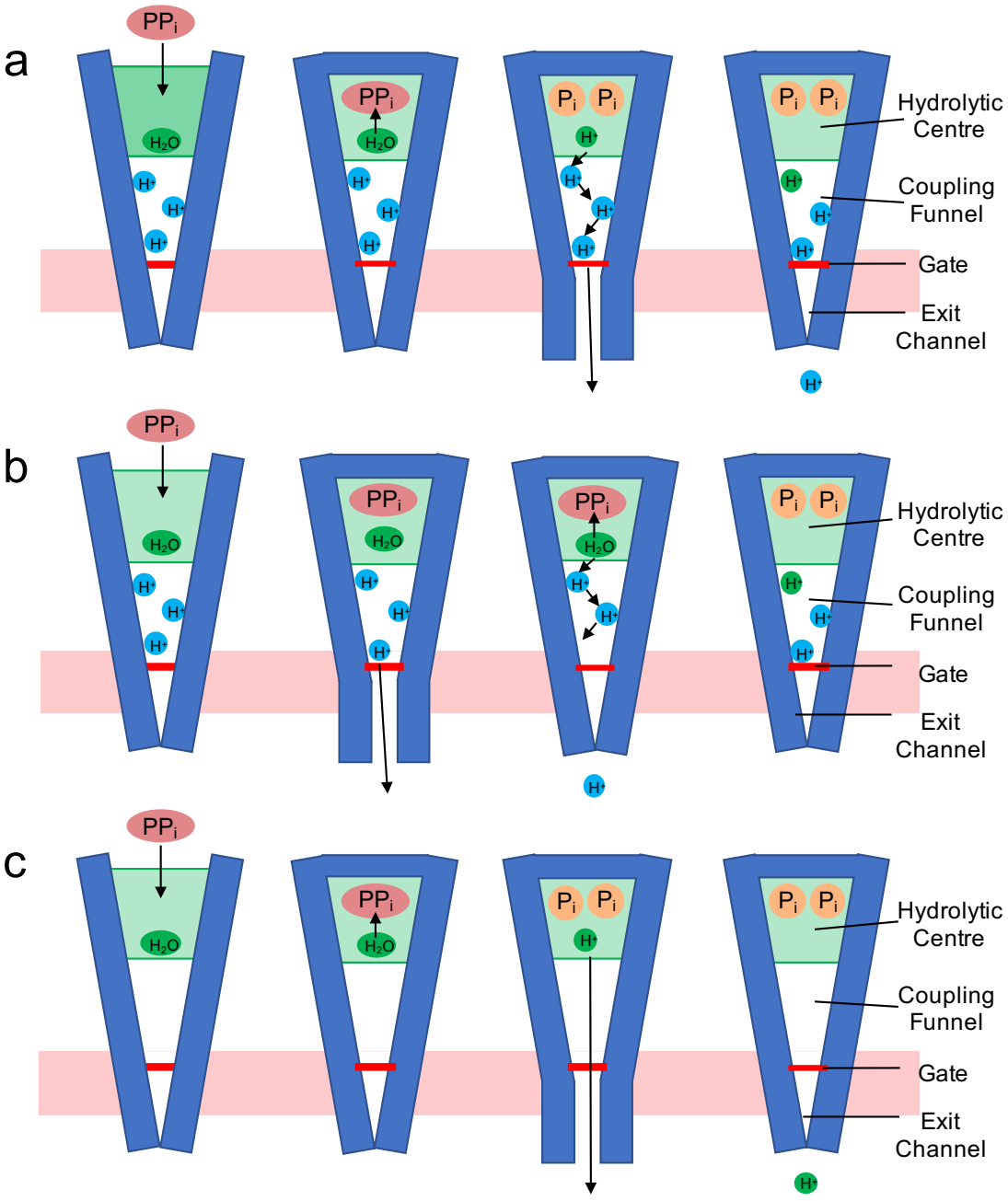


Figure 1.11: Proposed Mechanisms of the Membrane Pyrophosphatase Catalytic Cycle.
Overview of the three proposed mechanisms of the M-PPase catalytic cycle. (a) Mechanism proposed by Lin et al., 2012 showing hydrolysis before pumping. (b) Binding-change mechanism proposed by Kellosalo et al. 2012, showing ion pumping before hydrolysis. Both (a) and (b) utilising a Grotthuss chain mechanism for translocation of ions along the coupling funnel. (c) Direct-coupling method proposed by Baykov et al. 2013, in which the proton abstracted from the hydrolysis of PP_i is pumped directly without entering a Grotthuss chain.

1.6 Aims and Objectives of the Thesis

The first aim of my thesis was to try and crystallise the M-PPase from *Thermotoga maritima* in as many different catalytic states as possible. These structures will be used alongside data on conformational changes, obtained via tryptophan fluorescence, and

proton pumping studies carried out using the SURFE2R technique to generate and update the complete model of the catalytic cycle. This will be used to try and understand the specific mechanisms through which these proteins hydrolyse PP_i, transport cations and couple these processes together. The second aim was to express, purify and crystallise the M-PPase from *Pyrobaculum aerophilum* and attempt to solve the structure of this protein using X-ray crystallography. In addition, mutagenesis experiments will be carried out to try and better understand the K^+-independence mechanism of this protein and to investigate the role of the C-terminus in protein structure and function. This information will further our knowledge of transporters, particularly M-PPases, and allow for the comparison of the different types of M-PPases (Na^+ or H^+, K^+-dependent/independent) to help understand the structural and functional differences between these mechanisms. This information also assisted in current drug discovery and drug design programmes, aiming to combat pathogens such as the Malaria-causing *Plasmodium* parasites and pathogenic prokaryotes that contain M-PPases.

Chapter 2:

Materials and Methods

2.1. Expression and Purification of M-PPases

2.1.1. Cloning of TmPPase and PaPPase Genes into Expression Vector

The yeast expression plasmid pRS1024 was used in all yeast expression experiments. The plasmid containing TmPPase (pMP649-Tm) was provided by Juho Kellosalo (University of Helsinki) and had been used previously to express TmPPase (Kellosalo *et al.*, 2011). The TmPPase gene contained several thermostabilising mutations, V353L and S395G (loop 8-9 and 9-10, respectively), and the plasmid contained an N-terminal 6-His tag (Figure 2.1).

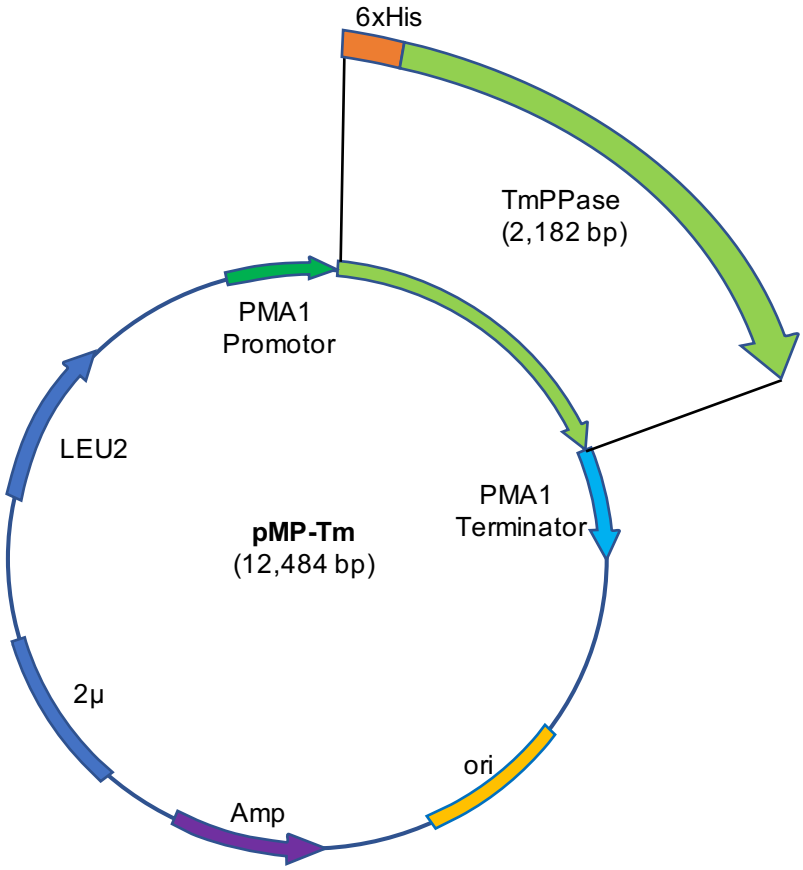


Figure 2.1: Schematic Overview of TmPPase-containing Expression Plasmid.
Overview of the pRS1024 yeast expression plasmid containing the 6-His tagged TmPPase, the PMA1 constitutive promotor and terminator sequences, LEU2 yeast selective marker and Ampicillin resistance gene used for bacterial selection.

Four different PaPPase constructs were generated (Pa A - Pa D) (Figure 2.2)(Table 2.1), two containing the signal peptide from *Trypanosoma cruzi* and two containing a GG-linker between the His-tag and gene sequence. The wild-type PaPPase gene was PCR amplified using primers containing the SalI and XbaI restriction sites (Table 2.2). Primers 91 and 93 were used to PCR amplify the PaPPase gene with the GG linker between the

His-tag and gene sequence and primers 90 and 93 were used to PCR amplify the gene without the GG linker (Table 2.2). Each set of primers was used on two different plasmid sources of the PaPPase gene. One plasmid contained the *T. cruzi* signal sequence upstream of the gene and the other did not. This signal sequence had previously been shown to improve the yield of protozoan parasite M-PPases expressed in yeast (Kellosalo *et al.*, 2011). The yield likely improved due to better localisation of the protein to the outer membrane, preventing it from saturating organelle membranes within the yeast expression host. The pRS1024 expression vector was linearised with XhoI and SpeI (isoschizomers of SalI and XbaI, respectively) restriction enzymes (Table 2.2) and the gene insert was ligated into the linearised vector using T4 DNA ligase and confirmed via colony PCR and sequence analysis using primers complimentary to the gene sequence (Table 2.2, primers 104-108).

Table 2.1: PaPPase Construct Overview.
List of all four constructs generated as part of the expression trials of Pyrobaculum aerophilum membrane pyrophosphatase, including presence or absence of signal peptide or GG-linker.

Name of construct	***Trypanosoma cruzi* signal peptide**	**GG-linker**
Pa A	No	No
Pa B	No	Yes
Pa C	Yes	No
Pa D	Yes	Yes

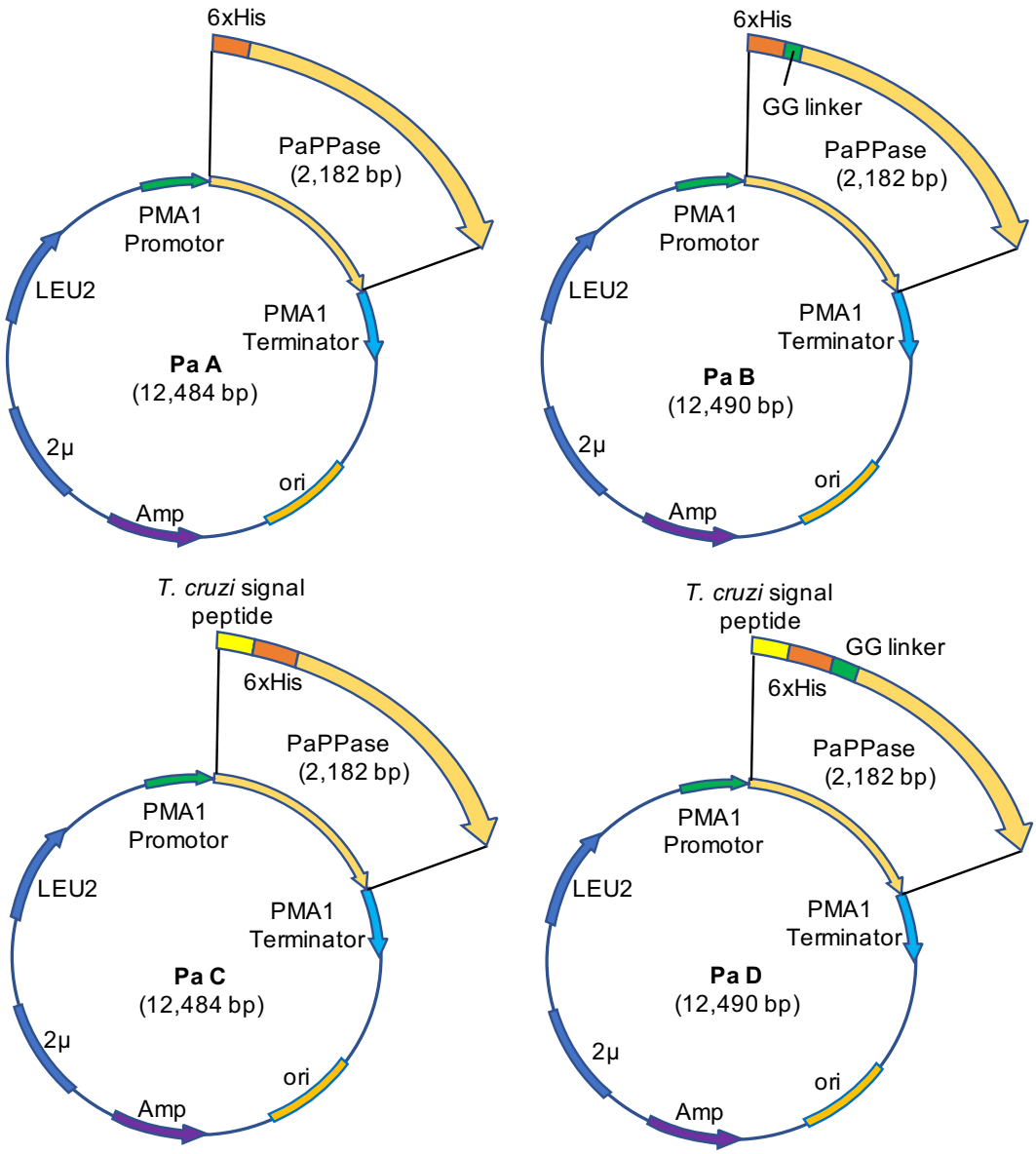


Figure 2.2: Schematic Overview of PaPPase-containing Expression Plasmids.
Overview of the four tested PaPPase constructs (Pa A – D), containing different combinations of a glycine-glycine linker between the His-tag and gene and a signal peptide from Trypanosoma cruzi upstream of the His-tag. All constructs were based on the pRS1024 yeast expression plasmid containing the PMA1 constitutive promotor and terminator sequences, LEU2 yeast selective marker and Ampicillin resistance gene used for bacterial selection.

Mutants of PaPPase were generated using the Q5 site-directed mutagenesis kit (NEB) to test the potassium-dependence. Primer pairs were prepared following the manufacturer's protocols (NEB) (Table 2.2, primers 119-124) and contained the mutation of interest. These were used in a standard PCR to amplify the entire plasmid, mutating the desired nucleotide in the process. The resulting constructs were digested with DpnI to remove remaining template DNA and the presence of the mutation was confirmed via sequence analysis.

Table 2.2: Cloning and Mutagenesis Primer List.

List of primers used for cloning and sequencing of PaPPase constructs. Annealing temperatures are based on manufacturer's guidelines for Q5 DNA polymerase (NEB). Nucleotide sequences shown in lower case represent mutations to the wild-type gene sequence and underlined regions represent restriction sites. All sequences are shown 5' to 3'.

Primer No.	Primer sequence	Restriction sites	direction	Tm (˚C)
Cloning primers				
90	ttt ttt GTC GAC ATG CAT CAC CAT CAC CAT CAC AAT ATG ATA AGC TAT GCC TTA CTA GG	SalI	F	59 (Q5)
91	ttt ttt GTC GAC ATG CAT CAC CAT CAC CAT CAC GGT GGA AAT ATG ATA AGC TAT GCC TTA CTA GG	SalI	F	59 (Q5)
93	ttt ttt TCT AGA TCA GAA AGG CAA TAG ACC TG	XbaI	R	59 (Q5)
Sequencing primers				
104	CG CAT TTT CTC GTT ACC A	N/A	R	N/A
105	AAT ATG ATA AGC TAT GCC	N/A	F	N/A
106	A GAT TTA GTT GGG AAG GT	N/A	F	N/A
107	ACT GTG CCT GTA AGC GGC	N/A	F	N/A
108	ATT GTC CTC CCC TTA ATA	N/A	F	N/A
Q5 site-directed mutagenesis primers				
119	C AAT ACC ACA gca GCC ACT ACT AAG GG	N/A	F	64
120	CC GAC GGA GTC CAG TAC A	N/A	R	64
121	A AAA GCC ACT gct AAG GGA TAT GC	N/A	F	65
122	GT GGT ATT GCC GAC GGA G	N/A	R	65
123	act gct AAG GGA TAT GCA ATA GCC AG	N/A	F	63
124	ggc tgc TGT GGT ATT GCC GAC GGA	N/A	R	63
InFusion cloning primers				
127	GCT GCC GCG CGG CAC	N/A	R	N/A
128	TCT AGT CGA GGG AAT TGT CAT T	N/A	F	N/A
129	CAC GGC GCG CCG TCG CATATGATAAGCTATGCCTTA	N/A	F	N/A
130	ATT CCC TCG ACT AGA tca CGC TAT ATT CGT AGA CAC GAT	N/A	R	N/A

C-terminal truncation mutants were generated using the In-Fusion cloning technique with primers specific to the N-terminus and 39 nucleotides upstream of the C-terminus region of the wild-type PaPPase gene (Table 2.2, primers 129-130). Each primer had a 15 bp overlap with the termini of the pRS1024 expression plasmid to allow for homologous recombination of the mutated gene into the plasmid. The pRS1024 expression plasmid was linearised using PCR with primers specific for each side of the multiple cloning site (Table 2.2, primers 127-128). Presence of the mutation was confirmed via sequence analysis of the constructs.

2.1.2. Transformation of Expression Vector into Yeast

S. cerevisiae BJ1991 yeast cells were first spread onto a yeast extract peptone dextrose (YPD) media plate and grown for 2 days at 28 ˚C. Individual colonies were picked and cultured in 7.5 ml YPD media overnight. 1 ml of overnight yeast culture per transformation reaction was pelleted in a micro-centrifuge and resuspended in 0.15 ml PLATE solution (40 % PEG 3350, 0.1 M LiAc, 10 mM Tris pH 7.5 and 1 mM EDTA). 10 µg of denatured fish sperm DNA (heated to 100˚C and cooled to room temperature) was added to each sample to reduce the degradation of the target DNA. 500 ng of target plasmid DNA was then added to each sample along with 10 µl DMSO, mixing briefly with a vortex mixer between each step. Samples were incubated at 28 ˚C for 15 minutes and then incubated at 42 ˚C for 20 minutes to heat-shock the yeast cells. Samples were then pelleted, the supernatant discarded and the cells were gently resuspended in 100 µl TE buffer (10 mM Tris pH 7.5 and 1 mM EDTA). The resulting mixture was spread onto plates containing Synthetic Complete Dropout, minus leucine (SCD-Leu) media and incubated for at least two days at 28 ˚C until colonies formed.

2.1.3. Expression of TmPPase

Yeast starter cultures were prepared from positive transformation colonies of the TmPPase-pRS1024 plasmid in Synthetic Complete Dropout minus leucine (SCD-Leu) media. Standard expression experiments were carried out in 10 L batches, with starter cultures consisting of 3 x 850 ml SCD-Leu media. These were inoculated with several loopfuls of yeast cells from the transformation plates and incubated at 30 ˚C for 24 hours, shaking at 200 rpm to maintain the cells in suspension. Starter culture was then inoculated into 10 x 750 ml Yeast extract Peptone Dextrose (YPD) media (250 ml culture per flask) and the protein was expressed by incubation at 30 ˚C for 8 hours. Following expression, cells were harvested and spun at 2,000 x *g* for 5 minutes (JLA 8.1000 rotor, Beckman). Pellets were washed via alternating resuspension in $_{dd}H_2O$ and spun at 2,000 x *g* for 5 minutes. This was repeated several times to ensure the pellets were completely washed. Pellets were finally collected via removal of all liquid, weighed and then stored at -80 ˚C until required.

2.1.4. Expression of PaPPase

Small-scale PaPPase expression trials were carried out by inoculating 187.5 ml of either SCD-Leu media or YPD media with 62.5 ml of starter culture containing each PaPPase

construct. These were incubated for 8 hours and the OD_{600} was monitored using a spectrophotometer (WPA Biowave II). Cells were harvested, washed and the pellets were stored at –80 ˚C, as described above. Larger-scale PaPPase expression trials were carried out following a modified version of the standard expression protocol for TmPPase (section 2.1.3) in which the expression cultures were incubated for 10 hours instead of 8 hours. OD_{600} was monitored using a spectrophotometer and cells were harvested, washed and the pellets were stored at -80 ˚C, as described above.

2.1.5. Membrane Extraction

Cell pellets from the expression experiments were thawed and resuspended in 1 ml of Buffer A (0.2 M Tris, pH 7.5, 40 % glycerol, 10 mM EDTA, 2 mM DTT and 0.2 mM PMSF) per 2 g of dry cell weight. Cells were lysed using 0.5 mm glass beads either in 50 ml Falcon tubes (< 20 ml resuspended cells) or in the bead-beater (Biospec products) (> 20 ml resuspended cells). 1.5 g of ice-cold glass beads were added per 1 ml of resuspended cells and the Falcon tube or bead-beater chamber was completely filled using buffer B (10 mM Tris pH 7.5, 10 % glycerol, 5 mM EDTA and 1 mM DTT). Cells were lysed via 12 x 1 minute mixing with the glass beads, interspaced by 1 minute cool-down on ice. Care was taken to ensure the samples remained cold at all times, to prevent damaging the proteins. The lysate and bead sample was spun at 4,000 x *g* to remove beads and cell debris. The supernatant was collected and spun at 42,500 rpm for 45 minutes using a pre-cooled Ti-45 rotor (Beckman) to pellet the membranes. Membrane pellets were collected and resuspended in buffer C (50 mM MES-NaOH pH 6.5, 20 % glycerol, 50 mM KCl, 5 mM MgCl2, 1.33 mM DTT, 0.336 mM PMSF and 2 µg/ml Pepstatin A) using a glass dounce homogeniser at room temperature.

2.1.6. Bradford Assay

The total protein concentration of the extracted membranes was calculated using the Bradford assay (Bradford, 1976), using bovine gamma globulin (BgG) as standard samples and a buffer control consisting of dilutions of buffer C. Extracted membrane samples were prepared at 1/10, 1/50, 1/100, 1/500, 1/1000 and 1/5000 in $_{dd}H_2O$. 20 µl of each sample was added to a 96-well plate in triplicate followed by 180 µl of Bradford reagent. The plate was incubated at room temperature for 20 minutes at which time the absorbance was measured at 600 nm. The total protein concentration was calculated, factoring in the buffer control and ensuring the sample was within the standard range.

Once the total protein concentration of the membranes had been determined, aliquots were flash-frozen in liquid nitrogen and stored at -80 ˚C.

2.1.7. Detergent Solubilisation of M-PPases

2.1.7.1. Standard TmPPase Solubilisation Protocol

Extracted membrane samples were diluted to 7.2 mg/ml total protein concentration using buffer C. Solubilisation was carried out in batches of 30 ml diluted membranes. 100 µl of 100 mM Na_2PP_i was added to each 30-ml batch of diluted membrane sample to ensure the sample remained stable at high temperature. Samples were pre-incubated at 75 ˚C along with 10 ml of solubilisation buffer (50 mM MES-NaOH pH 6.5, 20 % glycerol and 5.34 % DDM) per 30 ml diluted membranes, for 15 minutes. The ratio of extracted membranes to detergent was 1:3 to allow for efficient solubilisation to occur (López-Marqués *et al.*, 2005). After incubation, both samples were combined and incubated at 75 ˚C for 1.5 hours, either shaking at 150 rpm or with manual inversion every 5 minutes. The resulting mix was spun at 4,500 x *g* for 5 minutes at room temperature to pellet the insoluble fraction and remove any remaining cell debris. The supernatant was then ready for column binding.

2.1.7.2. Small-scale PaPPase Solubilisation Screen

A small-scale solubilisation test was carried out for PaPPase using a commercial screen of 20 detergents (Table 2.3). For each detergent, 200 µl of diluted membrane sample at 5 mg/ml was used. This was mixed with 66.6 µl of solubilisation buffer with 1 - 50 x CMC of each detergent and incubated at 70 ˚C for 1 hour, shaking at 1,000 rpm. 220 µl of each sample was then spun at 100,000 x *g* and 4 ˚C for 1 hour using a TLA-100 rotor (Beckman). Each pellet was resuspended in 220 µl resuspension buffer (50 mM MES pH 6.5 and 50 mM KCl) and used for SDS-PAGE analysis and activity assay measurements.

Table 2.3: List of Detergents used for Solubilisation Tests.
List of all detergents used during the small-scale solubilisation test of Pyrobaculum aerophilum Membrane Pyrophosphatase along with associated Critical Micellar Concentrations (CMCs) and acyl chain length.

Name of detergent	Abbreviated name	Chain length	CMC (%)
7-Cyclohexyl-1-Heptyl-β-D-Maltoside	CYMAL-7	7C + ring	0.0099
n-Octyl−β−D-Thiomaltopyranoside	OTM	8C + S	0.53
n-Nonyl−β−D-Thiomaltopyranoside	NTM	9C + S	0.15
n-Decyl−β−D-Thiomaltopyranoside	DTM	10C + S	0.045
n-Undecyl−β−D-Maltopyranoside	UM	11C	0.029
n-Undecyl−β−D-Thiomaltopyranoside	UTM	11C + S	0.011
n-Dodecyl−β−D-Maltopyranoside	DDM	12C	0.0087
n-Dodecyl−β−D-Thiomaltopyranoside	DDTM	12C + S	0.0026
n-Tridecyl−β−D-Maltopyranoside	TriDM	13C	0.0017
n-Tetradecyl−β−D-Maltopyranoside	TetraDM	14C	0.00054
2,6-Dimethyl-4-Heptyl−β−D-Maltopyranoside	2,6-Dimethyl-4-Heptyl−β−D-M	9C (forked tail)	1.2
3-Cyclohexyl-1-Propyl-β-D-Glucoside	CYGLU-3	3C + ring	0.86
n-Nonyl−β−D-Glucopyranoside	NG	9C	0.2
Methyl-6-O-(N-Heptylcarbamoyl)-α-D-Glucopyranoside	Anameg-7	7C	0.65
β-D-Fructopyranosyl-α-D-Glucopyranoside Monododecanoate	Sucrose monododeconoate	X	0.016
n-Heptyl−β−D-Thioglucopyranoside	HTG	7C + S	0.85
Tetraethylene Glycol Monooctyl Ether	C8E4	X	0.25
Pentaethylene Glycol Monodecyl Ether	C10E5	X	0.031
Hexaethylene Glycol Monooctyl Ether	C8E6	X	0.39
Octaethylene Glycol Monododecyl Ether	C12E8	X	0.0048

2.1.7.3. Large-scale PaPPase Solubilisation Test

Large-scale solubilisation tests of PaPPase followed a modified version of the TmPPase solubilisation protocol (Section 2.1.7.1) in which DM, UM and DDM were used in place of just DDM. During solubilisation, 100 µl of 100 mM Na_2PP_i was substituted for 100 µl of 100 mM Na_2IDP or 100 µl water. All other steps were the same as the standard TmPPase protocol (Section 2.1.7.1).

2.1.8. Ni-NTA Purification of M-PPases

2.1.8.1. Standard TmPPase Purification Protocol

Nickel-Nitrilotiacetic Acid (Ni-NTA) resin columns were prepared by adding 0.5 ml of 80 % Ni-NTA matrix slurry to every 40 ml of solubilised membrane sample. The slurry was washed with 10 x volume of buffer C supplemented with 0.2 % DDM, spun at 1,800 x *g* and 4 ˚C for 3 minutes, after which the supernatant was discarded. For every 40 ml of solubilised membrane sample, 32 ml of buffer C, 6 ml of 3.5 M KCl and 0.5 ml of

washed Ni-NTA beads were combined and incubated for 1.5 hours at 40˚C, shaking at 100-150 rpm to ensure the beads remained in solution.

Following incubation, the samples were added to 20 ml columns. The initial flow-through sample was collected for use in SDS-PAGE. The column was washed with 2 x column volume (CV) of wash buffer (50 mM MES-NaOH pH 6.5, 20 % glycerol, 50 mM KCl, 5 mM $MgCl_2$, 1 mM Na_2P_i, 30 mM imidazole, 0.02 % DDM, 1 mM DTT, 0.2 mM PMSF and 2 µg/µl pepstatin A) followed by one wash containing 2 x CV of elution buffer (50 mM MES-NaOH, pH 6.5, 3.5 % glycerol, 50 mM KCl, 5 mM $MgCl_2$, 1 mM Na_2P_i, 400 mM imidazole, 0.02 % DDM, 1 mM DTT, 0.2 mM PMSF and 2 µg/ml pepstatin A). The resulting samples were collected and used for SDS-PAGE analysis and concentration steps.

2.1.8.2. PaPPase Purification

During purification trials of PaPPase, an imidazole gradient was used to test the different PaPPase constructs following a modified version of the standard TmPPase purification protocol (Section 2.1.8.1). Samples and columns were prepared following the standard protocol, but a series of wash buffers were prepared covering an imidazole gradient to final concentrations of 10, 20, 30, 40, 50, 100, 200 and 300 mM imidazole. Larger-scale purification of PaPPase was carried out following the standard TmPPase purification protocol except the concentration of imidazole in the wash buffer was 20 mM instead of 30 mM.

2.1.8.3. On-column Detergent Exchange

Detergents were exchanged on the Ni-NTA column, as required. PaPPase was exchanged from DDM into DM, OGNG, LMNG, OM, NM and NG by replacing DDM in the wash and elution buffers with 4 x CMC of each buffer. All downstream purification buffers then used the new detergent in place of DDM at 4 x CMC.

2.1.8.4. Protein Concentration and Buffer Exchange

Eluted samples from the Ni-NTA purification were concentrated via centrifugation of the sample through a 50 kDa vivaspin 2 (MWCO) protein concentrator (GE Healthcare), which allows the liquid to pass but not the protein or detergent micelles. The concentration of protein was determined from A280 readings using the DS11+ UV

spectrophotometer, blanked against the elution buffer and the protein was concentrated to between 10 and 20 mg/ml, suitable for downstream applications.

Following concentration of the protein sample, buffer exchange was carried out to remove the imidazole. This was achieved using the micro bio-spin 6 columns (Bio-Rad), which were preloaded with the M-PPase crystallisation buffer (3.5 % glycerol, 50 mM KCl, 20 mM MES-NaOH pH 6.5, 5 mM $MgCl_2$, 2 mM DTT, 4 x CMC detergent), determined from previous crystallisation experiments of TmPPase (Kellosalo *et al.*, 2013). To equilibrate the columns, 4 x 500 ml buffer was added to each and these were spun at 1,000 x *g* for 1 minute between each addition. Following this, 20 - 75 µl of concentrated protein was added and the column was spun at 1,000 x *g* for 4 minutes. The concentration of the buffer exchanged protein was determined by A280 method using the crystallisation buffer as a blank. The protein was then either directly used for experiments or was flash-frozen in liquid nitrogen and stored at -80 ˚C. If alternative buffers were required, these were substituted in place of the crystallisation buffer.

2.1.8.5. SDS-PAGE

SDS-PAGE was carried out to check the purity of protein samples following purification and for carrying out Western Blot analysis. Protein samples were prepared by adding 10 µl of 4 x SDS sample buffer (200 mM Tris-HCl pH 6.8, 80 mg/ml SDS, 40 % glycerol, 50 mM EDTA, 0.8 mg/ml bromophenol blue and 20 mM DTT) to each 30 µl protein sample. In the case of more concentrated samples, such as the extracted membrane samples, these were diluted 1:5 prior to addition of 4 x SDS sample buffer. Precast gels (Biorad) were used and these were prerun at 120 V for 10 minutes with 1 x SDS running buffer (25 mM Tris, 192 mM glycine and 0.1 % (w/v) SDS) to ensure even distribution of SDS throughout the gel. Samples were loaded into the gel alongside a protein ladder (Color protein standard, NEB) and BSA standards, where applicable. Gels were run at 150 V for around 45 minutes, or until the dye front reached the bottom of the gel. Gels were then washed in $_{dd}H_2O$ for 15 minutes on a shake platform, stained with quick Coomassie stain (Generon) for 30 minutes when bands became visible on the gel. Gels were destained by washing in $_{dd}H_2O$ for 1 hour on a shake platform and were imaged in the G:Box (Syngene) gel imaging system. Gels used for Western blot analysis were not washed or stained but were immediately blotted instead.

2.1.8.6. Western Blot

Western blot analysis was carried out to visualise specific protein bands using protein or tag-specific antibodies following SDS-PAGE analysis. SDS-PAGE was carried out as described above and the gel was transferred to a 0.2 µM PVDF membrane (Bio-Rad) using the Trans-blot Turbo transfer system (Bio-Rad) set to 2.5 A and 25 V for 10 minutes. Following transfer, membranes were blocked with TBS-T (137 mM NaCl, 2.7 mM KCl, 19 mM Tris and 0.1 % Tween 20) containing 3 % BSA for 1.5 hours, rolling at room temperature. Rabbit Anti-M-PPase antibody was added at a 1:10,000 dilution and the membrane was rolled for 1 hour at room temperature. The membrane was then washed three times with TBS-T buffer, rolling for 10 minutes during each wash and discarding buffer between. Fresh TBS-T buffer containing 1:10,000 dilution of the secondary antibody (goat anti-rabbit) was added and the membrane was incubated for 1 hour at room temperature on a rolling platform. Membranes were washed three times again with TBS-T for 10 minutes each and then 1.5 ml of Clarity Western ECL substrate (Bio-Rad) was added per membrane to develop the bands. Blots were imaged using the G:Box gel imaging system.

2.1.9. M-PPase Hydrolytic Activity Assay

2.1.9.1. Relipidation of Protein Samples

2.1.9.1.1. Standard TmPPase Relipidation Protocol

Prior to running the activity assay, protein samples must be relipidated to replace lipids lost during purification and so recover maximal activity. TmPPase samples were relipidated by mixing 40 µl of 30 mg/ml soybean lecithin dissolved in 20 mM Tris-HCl pH 8.0 and 1 mM DTT (sonicated until clear) with 22.5 µl of 20 % DDM. The solution was heated at 55 ˚C for 15 minutes, allowed to cool to room temperature before 35.5 µl of buffer C was added. The solution was mixed and 2 µl of concentrated TmPPase was added and the solution kept on ice prior to activity assay measurements.

2.1.9.1.2. PaPPase Relipidation Trial

Several relipidation methods were tested for PaPPase based on modified versions of the standard TmPPase relipidation protocol (Section 2.1.9.1.1). The first method followed the TmPPase protocol. The second method followed previous studies into PaPPase purification (Kellosalo *et al.*, 2011) and involved addition of 47 µM soybean lecithin to the protein sample, with 20 % DDM replaced with additional buffer C. The third method

also followed the standard TmPPase protocol except the 20 % DDM was replaced with additional buffer C, the soybean lecithin concentrations were kept the same as the TmPPase protocol.

2.1.9.2. Standard TmPPase Activity Assay Protocol

The main reagents were prepared in advance of the assay and consists of solution A (0.3 g ascorbic acid added to 10 ml ice-cold 0.5 M HCl) and solution B (70 mg Ammonium heptamolybdate added to 1 ml ice-cold H_2O). Both solutions are combined and kept on ice for at least 30 minutes, but no longer than 3 hours as the colour starts to develop even in the absence of protein. A sodium arsenite solution, which stops the colour change once added, was also prepared by dissolving 5 g of sodium arsenite and 5 g of trisodiumcitrate in 100 ml of H_2O. 5 ml of acetic acid was then mixed into the solution and this was taken to a maximum volume of 250 ml with H_2O. Care was also taken to ensure sufficient buffer was used to keep the detergent concentration of each sample above the critical micellar concentration (CMC) (the minimum concentration of detergent at which micelles will still form) usually aiming to keep it between 2 and 4 times above the CMC.

Initial M-PPase activity assays were carried out in standard 1 ml cuvettes. Each protein sample was prepared to a final volume of 240 µl in activity assay buffer (80 mM Tris-HCl pH 8.0, 5 mM MgCl2, 120 mM KCl, 20 mM NaCl, 1 mM NaF) and used 0.2 µl and 1 µl of relipidated protein. Phosphate standards were prepared using Na_2P_i and control samples were prepared using only activity assay buffer and these were run alongside protein samples. Solution A and solution B were mixed and kept on ice for at least 30 minutes but no longer than 3 hours. To carry out the experiment, each sample was placed into a 71 ˚C heat block in sequence, one tube every 15 seconds. After 5 minutes, 10 µl of 10 mM Na_2PP_i was added to each tube in sequence, one tube every 15 seconds, and incubated for 5 minutes at 71 ˚C. Samples were transferred to ice in sequence, one tube every 15 seconds, and incubated for 10 minutes. 300 µl of solutions A and B was added to each tube in sequence, every 15 seconds, and samples were kept on ice for 10 minutes. 450 µl of sodium arsenite solution was added to each tube in sequence, every 15 seconds and the samples were incubated at room temperature for 1 hour. Absorbance was measured at 860 nm and the total activity or specific activity was calculated.

2.1.9.3. High-throughput 96-well Plate Protocol

A new method for measuring the hydrolytic activity of the protein was used for later M-PPase activity assay measurements. This method followed the same principle as the standard TmPPase activity assay protocol but used a 96-well plate and 8-tube PCR strips. Each sample was prepared as before, but only 40 µl of protein sample in activity assay buffer was added to each tube in the PCR strip. The same sequential addition of Na_2PP_i, solution A and B, and sodium arsenite solution was carried out, but on one PCR strip at a time instead of individual tubes. Strips were added to a 71 ˚C heat block, one strip every 20 seconds, and these were incubated for 5 minutes. 10 µl of Na_2PP_i was added to each strip using a multichannel pipette in sequence, one strip every 20 seconds, and incubated for 5 minutes. Strips were transferred to ice, one strip every 20 seconds, and these were incubated on ice for 10 minutes. 60 µl of solution A and B was added to each strip using a multichannel pipette, one strip every 20 seconds, and incubated on ice for 10 minutes. 90 µl of sodium arsenite was added to each sample using a multichannel pipette, one strip every 20 seconds, and these were incubated at room temperature for 1 hour. Samples were transferred from the PCR strips to the 96 well plate and the absorbance was measured at 860 nm using a plate reader.

2.2. Structural Studies of M-PPases Using X-ray Crystallography

2.2.1. Preparation of Crystallisation Screens

2.2.1.1. Sparse Matrix Screens to Obtain Initial Hits

Purified protein samples were exchanged into crystallisation buffer (3.5 % glycerol, 50 mM KCl, 20 mM MES-NaOH pH 6.5, 5 mM $MgCl_2$, 2 mM DTT) using previously described methods (Section 2.1.8.4) and used for protein crystallisation trials. Several commercially available sparse matrix screens were used, including MemGold 1 and 2 alongside MemStart and MemSys screens. These have been designed to include a range of conditions commonly associated with crystallisation of membrane proteins. 50 µl of the well solution from each screen was transferred to 96-well crystallisation plates (2-drop MRC plates) and crystallisation plates were set up using the NT8 crystallisation robot (Formulatrix). The sitting drop vapour diffusion method was used and drop sizes were 0.1 µl protein mixed with 0.1 µl well solution. Plates were stored in the RI1000 plate imager and hotel (Formulatrix) and each drop was automatically imaged every few days with normal, UV-TPEF and SONICC imaging and the resulting images were viewed using the Rockmaker software (Formulatrix).

2.2.1.2. Optimised PaPPase Crystallisation Screen

An optimised crystallisation screen was set up, based on the best hits observed from previous crystallisation attempts of PaPPase (Kellosalo *et al.*, 2013). This screen was split into two sections, the first had a pH range from 5.5 to 8.5 and contrasted this against PEG 400 concentrations from 29 % to 39 %, all using 0.1 M KCl as the only salt. The other half of the screen kept the buffer concentrations constant at either 0.1 M MES-NaOH, pH 6.5 or 0.1 M Tris, pH 8 and altered the salt types and concentrations, including 0.01 M, 0.05 M and 0.2 M KCl, 0.1 M NaCl and combinations of $CaCl_2$, $LiSO_4$ and NaCl. This screen was designed using the Rockmaker software, prepared using the Formulator robot (Formulatrix) and was used in setting up crystallisation plates using the methods described above.

2.2.1.3. Detergent Trials

Many variations of the sparse matrix and optimisation screens were set up for PaPPase. Crystallisation plates were set up with protein concentrations ranging between 5 and 10 mg/ml. The protein was exchanged into DDM, DM, LMNG, OGNG, OM, NM and NG during purification (Section 2.1.8.3) and crystallisation plates were set up for each, following the previously described methods (Section 2.2.1.1). When lipids were tested, these were prepared prior to setting up the crystallisation plates and following the same lipid:protein ratios found in the relipidation protocol (Section 2.1.9.1). In addition, several additives were tested to try and lock the protein into various states of the catalytic cycle. These included the addition of $CaCl_2$ and the inhibitor imidodiphosphate (IDP), which were independently mixed with the protein sample to a final concentration of 2 mM prior to setting up the crystal plates.

2.2.2. Crystal Optimisation Methods

2.2.2.1. High Lipid and Detergent (HiLiDe) Optimisation

The "HiLiDe" (high lipid and detergent concentration) method (Gourdon *et al.*, 2011) was also used to try and optimise crystals of PaPPase. This technique differs from standard crystallisation as increasing concentrations of two different detergents are added to the crystallisation screen (Figure 2.3). This technique involved setting up 24-well plates of both hanging and sitting drop vapour diffusion plates. Plates were prepared manually by adding 500 µl of the well solution to each well. The drop size was increased to 1 µl

protein plus 1 µl well solution. Conditions were centred around the best crystallisation conditions observed during this study and involved increasing concentrations of PEG 400 and various concentrations of KCl and Tris-HCl buffer. Where protein lipidation was used, this was carried out as described above (Section 2.1.9.1) using Soybean Lecithin. Increasing concentrations of DDM, DM and NM between 0 and 5 % were tested, and different combinations of these detergents were used. Plates were manually visualised under a standard light microscope to observe any potential crystal growth.

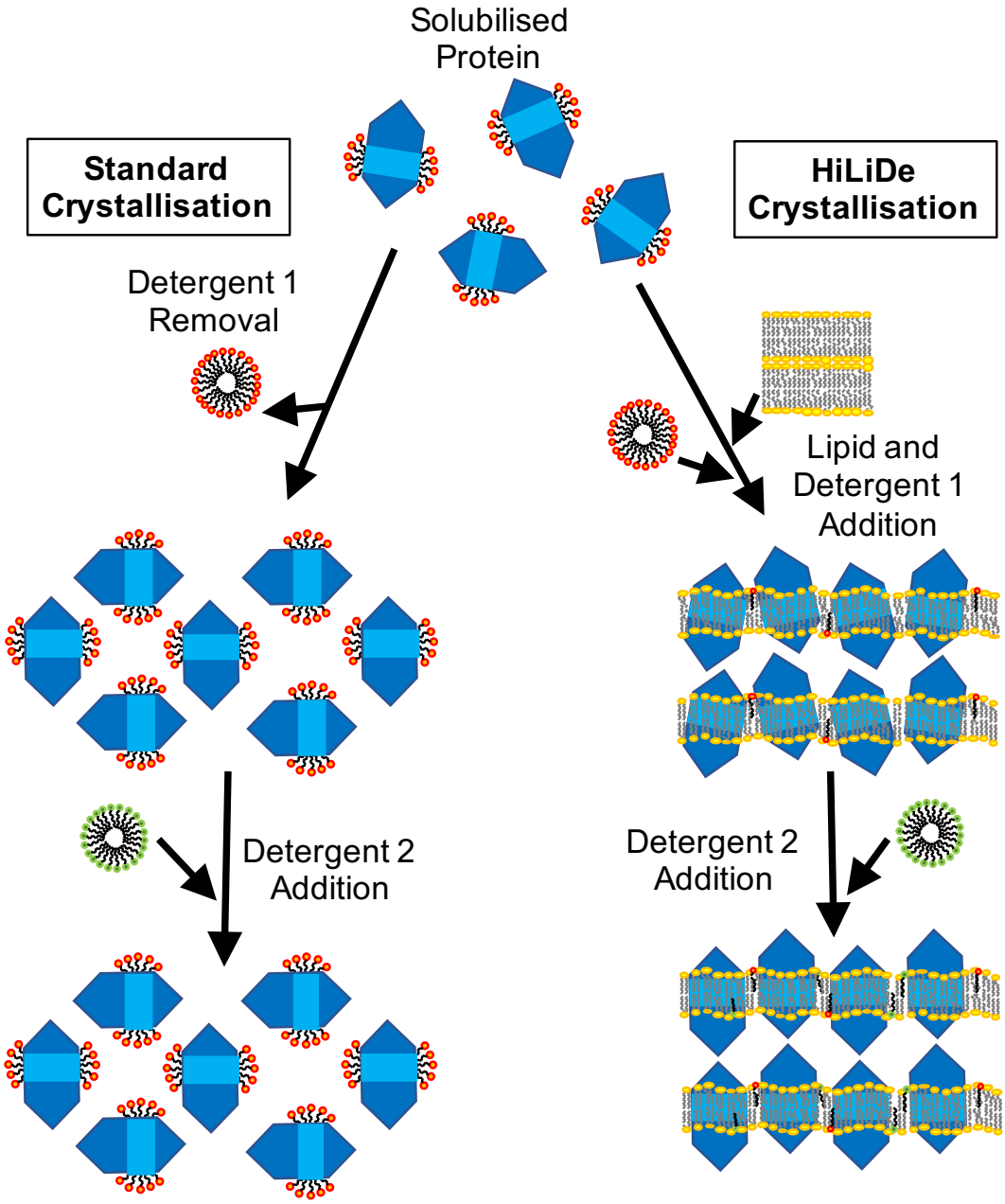


Figure 2.3: Schematic Overview of HiLiDe Crystal Optimisation Method.
Comparison between the standard crystallisation technique and HiLiDe optimisation method, involving the addition of lipids and two different detergents to increasingly high concentrations (~6 %). The addition of specific concentrations of each detergents and lipid can form stable type-2 crystals that may improve diffraction. Figure redrawn from Gourdon et al. (2011), Figure 1.

2.2.2.2. Crystal Dehydration Experiments

Dehydration experiments were carried out by transferring crystals to solution containing mother liquor plus either 10 % glycerol or increased PEG 400 concentrations (final concentration 5-10 % higher than initial crystallisation conditions) and monitoring plates for signs of additional crystal growth. Another method was to add increased glycerol concentrations (to a final concentration of 10 %) to the well solution, resealing the plate and monitoring crystal development over the course of several days to see if this improved crystal quality and diffraction. Additional methods of dehydration involved exposing the loops to the air during crystal picking or manually removing the excess liquid from the loop using blotting paper. This allowed for partial evaporation of the excess solution in the well and created a dehydrating effect on the crystals in the loop.

2.2.3. Crystal Harvesting & Data Collection

Crystal harvesting was carried out by manually picking crystals with either nylon, mesh or litholoops, appropriately sized to match each crystal, and these were immediately flash frozen in liquid nitrogen. Harvesting was carried out at 20 ˚C to limit unwanted evaporation of the drop solution. Crystals were either sent to beamline i04 or i24 at Diamond Light Source (DLS) or to beamlines ID23-1 and MASSIF-1 at the European Synchrotron Radiation Facility (ESRF) for data collection.

2.2.4. Data Processing and Structural Refinement

The highest resolution data sets from the synchrotrons were processed using a range of different software packages. Initial processing of PaPPase diffraction images and generation of intensities was carried out in XDS (Kabsch, 2010). Pointless (P. Evans, 2006) and Aimless (P. R. Evans & Murshudov, 2013), from the CCP4 software package (Winn *et al.*, 2011), were used to scale and merge the data. For the TmPPase structural studies, X-ray diffraction data was supplied by Juho Kellosalo (University of Helsinki) and initial processing had been carried out in XDS. One set had been acquired on beamline i04 at Diamond light source, based on crystals grown in 32 % PEG 400, 0.1 M Tris pH 8.5, 0.1 M $MgCl_2$, 0.1 M NaCl and 2 mM DTT. The other set was from data collected for a previous publication (Kellosalo *et al.*, 2012) wherein the data had been used for phasing other structures of TmPPase but had never been fully solved.

Molecular replacement was carried out for PaPPase using a modified search model based on the 2.6 Å resolution TmPPase resting state structure (PDB ID: 4AV3). This structure had been used to generate a homology model using the sequence of PaPPase. The TmPPase structures also used a modified version of the 2.6 Å model of TmPPase (PDB ID: 4AV3) as a search model, in which the metal ions and water molecules had been removed to prevent model bias.

Initial rounds of jelly body refinement were carried out in Refmac (CCP4), but refinement of the models was predominantly carried out using the phenix.refine (Afonine *et al.*, 2012) program in the Phenix software package with manual model alterations made in COOT (Emsley *et al.*, 2010). Refinement was carried out using reference model restraints from the highest resolution VrPPase structure at 2.35 Å (PDB ID: 4A01) and secondary structure restraints. These were used alongside torsion-angle NCS restraints and tighter stereo-chemical weighting to maintain realistic geometry and to prevent over-fitting the data at these resolutions.

The combined PaPPase dataset was generated from 7 individual datasets. These were individually processed in XDS, using only the first few hundred images from each dataset. All datasets were in the same space group, had less than a 2 % deviation in unit cell parameters and all had positive density (F_o-F_c) at the expected position of the Mg_5IDP complex, confirming the structures were in the same catalytic state. The individual datasets were merged and scaled using XSCALE and the resulting file was used for molecular replacement and refinement as described above.

2.2.5. Comparisons of M-PPase Structures

2.2.5.1. Structural Comparisons in PyMol and Coot

Structural analysis was carried out on all available structures, including the previously solved structures: TmPPase:CaMg (PDB ID: 4AV3), TmPPase:Mg_4P_{i2} (PDB ID: 4AV6) and VrPPase:Mg_5IDP (PDB ID: 4A01) as well as the structures I solved: TmPPase:Mg_5IDP (PDB ID: 5LZQ) and TmPPase:$MgWO_4$ (PDB ID: 5LZR) alongside those from our collaborators lab of VrPPase:Mg_3P_i. Structures were visualised, superimposed and analysis was carried out in COOT (Emsley *et al.*, 2010) and Pymol. Structural figures were prepared and rendered in Pymol.

2.2.5.2. Hydrogen Bond plot (HBplot)

The web server HBplot (Hydrogen Bond plot) (Bikadi *et al.*, 2007), was used to analyse the hydrogen bonding patterns of structures in the PDB format. HBplot generates a spreadsheet of the hydrogen bond network of the structure along with distances. The default settings of the server allowed for an overestimation of the number of hydrogen bonds as some of the search parameters for bonding distances can be unrealistically high (~ 4 Å). Therefore, the datasets were processed manually, removing outliers beyond the expected distances of hydrogen bonds and assigning the type of helix at each position (i.e. α, π, 3_{10}). The data were then compared and contrasted between the different catalytic states to observe how the hydrogen bond network changed across the catalytic cycle.

2.2.5.3. BENDIX

Bendix (Dahl *et al.*, 2012), a plugin for the Visual Molecular Dynamics (VMD) software (Humphrey *et al.*, 1996) was used to analyse regions of secondary structure, notably helices, from PDB models and measures the degree of curvature. This was outputted as a visual representation with a list of the angles of helices in degrees. This was carried out for all available structures and the information was manually checked, with comparisons made between specific helices of each structure, notably helices 6, 12 and 16 as these contains many of the important active site residues.

2.2.5.4. Valence Test

A valence test (Nayal & Di Cera, 1996) was carried out on a potential sodium ion via submission of the model file in PDB format to the WASP (Water Screening program) that analyses properties of placed water molecules and metal ions, scoring them on likelihood of being correct based on valences from the electron density. (Note: This server has now been superseded by the CheckMyMetal server (Zheng *et al.*, 2017), which carries out the same calculations. The original WASP server is no longer accessible)

2.3. Biochemical Studies of M-PPases

2.3.1. Tryptophan Fluorescence Experiments

By subjecting protein samples to specific wavelengths of light it is possible to excite the tryptophan residues so that they emit light at wavelengths detectable with a fluorimeter. This can be a useful way of detecting conformational changes in proteins, although the

signal generated depends on the number of tryptophan residues and their location compared to the mobile parts of the protein. Furthermore, this signal can be enhanced through the use of a compound called 8-anilino-1-napthalene sulphonate (ANS). This compound binds to hydrophobic regions of the protein, is excited by the emission wavelength of tryptophan fluorescence and will fluoresce at a different wavelength to that of tryptophan. This leads to an exponential increase in the signal from tryptophan fluorescence, which is useful in cases where the original tryptophan signal is weak.

2.3.1.1. Preparation of Samples for Tryptophan Fluorescence Experiments

The buffers used for this assay were based on the optimal conditions of the hydrolytic activity assay. This used Tris-HCl as a core buffer component with the appropriate metal salts (KCl, NaCl and $MgCl_2$) added for each M-PPase protein. Stock buffers were prepared for TmPPase (80 mM Tris-HCl, pH 8, 100 mM KCl, 5 mM $MgCl_2$ and 20 mM NaCl), VrPPase (80 mM Tris-HCl, pH 8, 100 mM KCl and 5 mM $MgCl_2$) and PaPPase (80 mM Tris-HCl, pH 8 and 5 mM $MgCl_2$). Protein samples were purified and exchanged into crystallisation buffer as described above.

2.3.1.2. Tryptophan Fluorescence Experimental Procedure

Protein samples were placed into a stirred cuvette in the fluorimeter (PTI) and left for 5 minutes to equilibrate prior to experiments. For each experimental run, 140 µg protein was used in the appropriate fluorescence buffer to a final volume of 975 µl. Total reaction volumes were 1000 µl, leaving 25 µl for the titrations, limiting the change in overall volume to 2 %. Titrations were carried out using increasing concentrations of stock solutions of inhibitors (IDP) as required (Table 2.4). Experiments were performed by adding the ligand, leaving the system to equilibrate for 30 seconds and then running an emission scan between 310 and 550 nm, with an excitation wavelength of 295 nm (tryptophan excitation wavelength). This scan covered any emissions of either tryptophan or ANS and was used to monitor changes to the ratios of these signals. Emission spectra were analysed in GraphPad Prism to obtain binding curves and Scatchard plots for each protein to observe the change in fluorescence with increasing concentrations of inhibitors or ligands.

Table 2.4: Titration Protocol for Tryptophan Fluorescence Experiments.
List of increasing concentrations and volumes of IDP needed for titration experiments of Membrane Pyrophosphatases for tryptophan fluorescence experiments. Final concentrations of IDP cover the biologically relevant ranges of IDP for membrane pyrophosphatases.

Stock [IDP] mM	Volume to add (µl)	Final [IDP] (µM)
0	0	0
1	1	1
1	1	2
1	3	5
1	5	10
10	1	20
10	3	50
10	5	100
100	1	200
100	3	500
100	5	1,000

2.3.2. SURFE^{2}R Experiments

The SURFE^{2}R N1 (Nanion Technologies, Germany) is a machine capable of measuring the difference in charge across a membrane, thus being capable of studying various ion translocation and pumping through channels and transporters. This technique is based on standard solid-state membrane techniques, combining a lipid-coated gold-surface sensor with a pump to provide immediate and real-time measurements of transport that can be measured directly as a change in current, a schematic overview of the technique is shown in Figure 2.4.

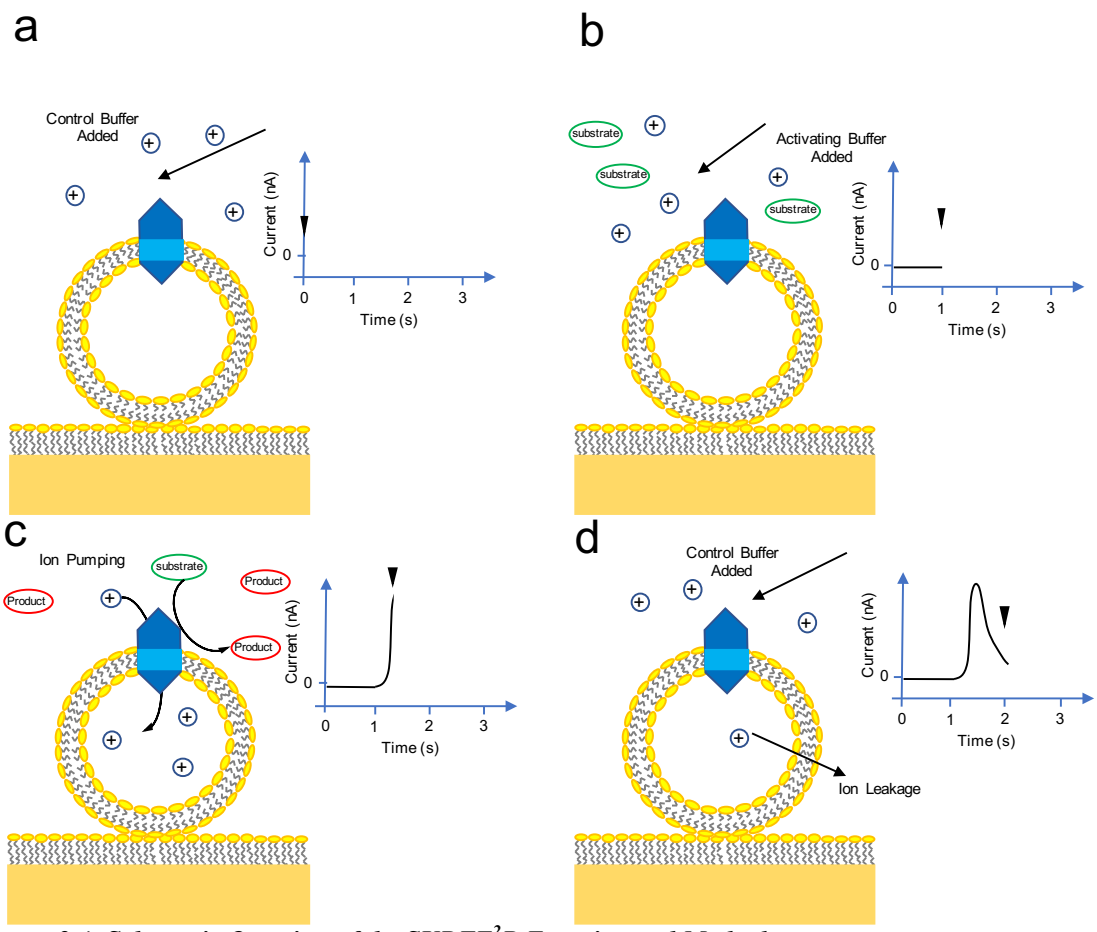


Figure 2.4: Schematic Overview of the SURFE^{2}R Experimental Method.
Overview of the experimental protocol for the SURFE^{2}R experiments showing a proteoliposome adhered to a membrane layer on a gold-surfaced sensor. (a) control buffer containing only the pumping substrate is added at time point zero to equilibrate the sensor (b) activating buffer containing the substrate or ligand of interest is added at one second time point. (c) after a short (ms) delay, pumping starts to occur into the liposome, showing as current on the graph (d) ion leakage from the membrane returns signal to baseline. Control buffer is added again at two second time point.

2.3.2.1. Reconstitution of VrPPase into Liposomes

SURFE^{2}R experiments were carried out on VrPPase protein generously provided by Prof. Yuh-Ju Sun (National Tsing Hua University, Taiwan). The protein was received frozen at 1 mg/ml in elution buffer (25 mM MOPS-KOH pH 7.0, 20 % glycerol, 400 mM KCl, 4 mM $MgCl_2$, 1mM PMSF and 250 mM imidazole).

To produce the liposomes, 120 mg/ml soybean L-α-phosphatidylcholine lipids were prepared. These were placed into an ice-water bath and sonicated at 30 % amplitude, 1 second pulse time for 1 minute interspaced with 1 minute rest on ice. This was repeated until the sample appeared clear. The protein sample was diluted to a final concentration of 0.05 mg/ml at 1 ml total volume using a reconstitution buffer (250 mM sorbitol, 1 mM $MgSO_4$, 100 µl EGTA, 2 mM DTT and 10 mM Tricine-Na pH 7.5). 15 µl of sonicated

liposome sample was combined with the 1 ml protein solution and prewashed Biobeads SM-2 (Bio-Rad) at a final concentration 0.25 mg/µl. The sample was mixed at 4˚C and 1,200 rpm for 1 hour to ensure the beads remained in suspension. Following this, the sample was desalted to remove residual glycerol and NaCl that might interfere with the experiments. This was carried out using the Micro Bio-spin 6 columns (Bio-Rad) equilibrated with the VrPPase reconstitution buffer (250 mM sorbitol, 1 mM $MgSO_4$, 100 µl EGTA, 2 mM DTT and 10 mM Tricine-Na pH 7.5). The resulting proteoliposome sample was divided into 20 µl aliquotes, flash frozen in liquid nitrogen and stored at -80 ˚C until required. Freeze-thawing of samples was strictly avoided to maintain the activity and stability of the liposomes.

2.3.2.2. Preparation of SURFE^{2}R Buffers

Three buffers were needed (A, B and C) for each SURFE^{2}R experiment (Table 2.5). Buffer A and B were the activating and non-activating buffers required for all experiments and, apart from the presence of substrate or a ligand in buffer A, are identical. Buffer C is used as a wash buffer and consists of only 0.1 M HEPES-KOH (pH 8) without any additional component such as salts or ligands. A concentrated stock of buffer C was used to generate all three buffers and was prepared at a working concentration of 100 mM and contained only the core buffer used for wash steps. A 10 x stock of buffer B was also prepared as this could be diluted down to a working stock of buffer B or, with the addition of substrate or ligand, could be made into a working stock of buffer A. This would help to limit the ionic variation between the two buffers and thus avoid interfering with the results.

Table 2.5: Buffer Compositions for SURFE^{2}R Experiments.
List of ingredients for all three buffers used during the SURFE2R experiments of proton pumping in the membrane pyrophosphatase of Vigna radiata. The 'x' denotes that concentrations were experiment-dependent.

Ingredient	Buffer A	Buffer B	Buffer C
HEPES-KOH, pH 8.0	0.1 M	0.1 M	0.1 M
$MgCl_2$	5 mM	5 mM	-
KCl	50 mM	50 mM	-
PP_i*	x	-	-
P_i**	-	x	-

* Both Na_4PP_i and K_4PP_i were tested
** Both Na_2HP_i and K_2HP_i were tested during the ionic balance experiments

2.3.2.3. SURFE^{2}R Sample and Sensor Preparation

SURFE^{2}R experiments were carried out using proprietary sensor chips (Nanion Technologies) and were prepared using standard protocols from the company. The chips were first washed and prepared via the addition of 50 µl sensor prep A2 (Nanion Technologies, proprietary buffer) directly to the gold surface of the sensor. This was incubated at room temperature for 15 minutes after which the liquid was removed. The sensors were washed with 3 x 100 µl isopropanol, removing the liquid in between each wash. The sensor surface was dried completely with nitrogen gas and 3 µl of the lipid-containing sensor prep B (Nanion Technologies, proprietary buffer) was added directly to the gold surface of the sensor, followed immediately by 50 µl of SURFE^{2}R buffer B. The sensors were incubated for 1 hour at 4 ˚C during which time the VrPPase proteoliposome sample was thawed on ice. This was diluted 1/5 into buffer B and sonicated at 20 % amplitude for 10 seconds (pulse time = 1 second) using a UP50H compact lab homogeniser (Hielscher). All remaining liquid was removed from the sensor after the 1 hour incubation and replaced with 50 µl of the sonicated proteoliposome sample. This was spun at 2,500 x *g* and 4 ˚C for 30 minutes and then incubated for at least 1 hour at 4 ˚C prior to use in experiments.

2.3.2.4. SURFE^{2}R Experimental Protocol

To carry out the experiments the sensor chips were placed into the SURFE^{2}R N1 along with buffers A, B and C. Buffer A was specific to each experiment, and contained

concentrations of K_4PP_i, Na_4PP_i or inhibitors at the desired concentrations for each experiment (5 – 100 μM). When gramicidin or CCCP were used, these were added to a final concentration of 10 μM in all three buffers and only added after initial measurements had been taken with substrate or inhibitor. Transfer of buffers was handled by the machine and the program was set up so that the surface of each sensor was initially washed twice with buffer C. Buffer B was injected onto the sensor for 1 second followed by rapid switching to buffer A for 1 second and then back to buffer B for 1 second. All measurements were taken twice, interspaced with two wash steps of buffer C and were measured as current (nA) over time (s).

Chapter 3: Expression and Purification of Membrane Pyrophosphatases

3.1. Introduction to Expression and Purification of M-PPases

The aim of this chapter was to express and purify the membrane pyrophosphatases from *Thermotoga maritima* (TmPPase) and *Pyrobaculum aerophilum* (PaPPase) for the purposes of downstream structural and functional characterisation. The TmPPase expressed and purified as part of this chapter was used to carry out functional characterisation, such as the Tryptophan fluorescence experiments, since structural studies were carried out on datasets collected by Juho Kellosalo (Helsinki). PaPPase was used for both structural studies and Tryptophan fluorescence experiments. Expression and purification of TmPPase had been carried out for previous studies and was expressed in yeast, following the "hot-solve" purification method (Kellosalo *et al.*, 2011; López-Marqués *et al.*, 2005). The protein is solubilised at 70 ˚C to remove contaminants, leaving only the thermostable TmPPase protein, which can remain active above these temperatures. This method was used to generate purified protein for the initial structural studies of TmPPase (Kellosalo *et al.*, 2012), which was one of the first M-PPase structures to be solved, alongside the VrPPase (Lin *et al.*, 2012). Ensuring the efficient expression and purification of these proteins was vital for downstream applications, such as X-ray crystallography, which requires large amounts of highly pure protein.

The "hot-solve" method was also used during the initial studies of PaPPase, in which the protein was expressed and purified alongside TmPPase (Kellosalo *et al.*, 2011). A range of different detergents were tested for hydrolytic activity and crystallisation of the protein, but the best combination yielded crystals that diffracted to around 5.8 Å, insufficient to gain specific mechanistic information on the protein. It was hoped that these crystals could be improved upon by optimisation of the expression and purification of PaPPase to improve the yield, alongside optimisation of the crystallisation conditions to improve crystal diffraction.

In addition to the optimisation of expression, purification and crystallisation of PaPPase, several mutants were also generated to test their biochemical effects on PaPPase. The first mutant was made to test the effect of removing 13 C-terminal residues on the crystallisation and diffraction of the resulting crystals. The second group of mutations was made to test the potassium-dependence of M-PPases. These mutants contained two key mutations found to be responsible for generating a potassium-independent mutant

from a potassium-dependent M-PPase in bacteria (Belogurov & Lahti, 2002). By reversing these mutations in PaPPase, it was hoped to generate a potassium-dependent variant to verify the findings of the previous study and provide more information on the mechanisms of potassium dependence. This could potentially allow for more accurate understanding on the reasons for having both variants in certain species of plants and parasites.

3.2. Expression and Purification of TmPPase

TmPPase can be expressed and purified exactly following previously published protocols (Kellosalo *et al.*, 2011). SDS-PAGE showed that the TmPPase samples from two Ni-NTA columns were pure (Figure 3.1). The main band was at the expected level for TmPPase, between the 58 and 80 kDa marker bands, with a single contaminating band level with the 245 kDa marker band, potentially representing aggregated or misfolded protein samples. The total yield of protein was determined from A280 measurements and was 1.18 mg/L culture prior to buffer exchange, and 0.9 mg protein/L yeast culture following buffer exchange. The yield is less than those obtained in previous studies (Kellosalo *et al.*, 2011) where a maximal yield of 2.5 mg/L was obtained. However, this can be explained through variability between spectrophotometers when measuring absorbance at 280 nm. The A280 reading of TmPPase purified in Helsinki was found to be 24 mg/ml, but the same sample measured in Leeds had a reading of only 12 mg/ml. Variability in measuring protein concentration could explain the difference in yield, since identical protocols were followed in both cases. The specific activity of the purified TmPPase protein was also found to be lower than previously reported (Kellosalo *et al.*, 2011), with a specific activity of 1.5 nmol P_i μg protein^{-1} min^{-1} as opposed to 150 nmol P_i μg protein^{-1} min^{-1} due to the presence of Ca^{2+} in all water and glassware in Leeds (see below).

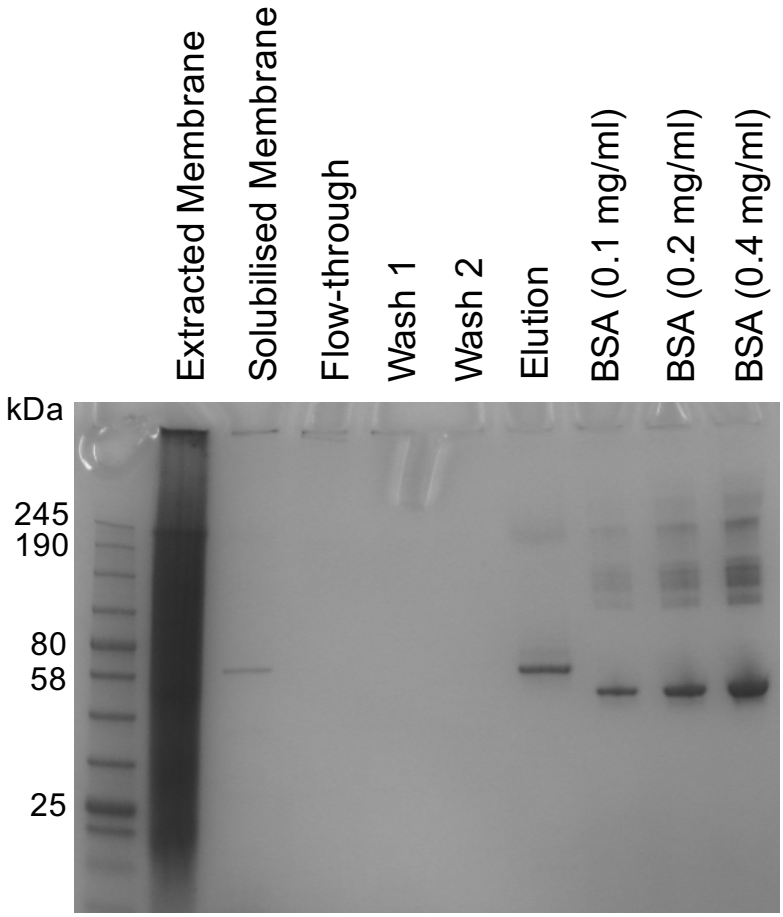


Figure 3.1: SDS-PAGE of TmPPase Purification.
Coomassie-stained SDS-PAGE of TmPPase samples obtained from yeast expression experiments. Including extracted membrane sample, solubilised membrane sample (DDM) as well as flow-through, wash and elution samples and BSA standards.

3.2.1. Overcoming Calcium Inhibition of Hydrolytic Activity

During purification of TmPPase, the hydrolytic activity was 10-100x lower than seen in previous studies (Kellosalo *et al.*, 2011). I investigated why this was occurring, and focussed on the complex interactions between the protein, pyrophosphate and magnesium ions. All of these are required for maximal activity, but the ratios can vary depending on relative concentrations of each component or indeed on the pH. Other divalent cations, such as calcium, have been shown to completely inhibit the activity of the protein (Maeshima, 1991). This makes the addition of calcium a useful technique for structural studies as it can lock the protein in the resting state (Kellosalo *et al.*, 2011) but can have significant effects on the activity of the protein during biochemical studies.

The first experiments looked at the role of magnesium-based compounds on the activity. Switching from $MgCl_2$ to $MgSO_4$ had no effect on the activity (Kellosalo *et al.*, 2011), nor did the addition of increasing concentration of $MgCl_2$ from 5 to 50 mM, with the latter having an inhibitory effect on the hydrolytic activity (Figure 3.2). This inhibition is likely due to precipitation of the magnesium-pyrophosphate complex at higher concentrations, as the solubility decreases significantly. This showed that the decreased activity compared with previous studies was not related to the source or concentration of magnesium.

Relipidation of the sample was previously shown to improve the activity of TmPPase by replacing lipids lost during purification (López-Marqués *et al.*, 2005). Relipidation alone was able to increase the activity 6-fold from the control sample (Figure 3.2). Adding 1 mM $MgCl_2$ to the relipidated sample led to a 3-fold increase in activity compared to the control sample, less than relipidation alone (Figure 3.2). The highest activity was seen when relipidation was combined with the addition of 160 μM EGTA, which showed a 17-fold increase in activity compared to the control sample (Figure 3.2). The use of EGTA, which is able to chelate free calcium with a much greater affinity than magnesium, implies that the primary cause of the decreased activity is calcium inhibition. Calcium is likely present in the water system in Leeds due to the geology of the area, since water in Helsinki is from granite (low pH and low Calcium) as opposed to the limestone of Leeds, which is high in calcium. The use of Milli-Q purified water and acid-washed glassware was insufficient to prevent loss of activity. Although these results were based on a single study without replication, the 16 x difference far exceed experimental error and a consistent hydrolytic activity was detected in future assays through the use of 160 μM EGTA.

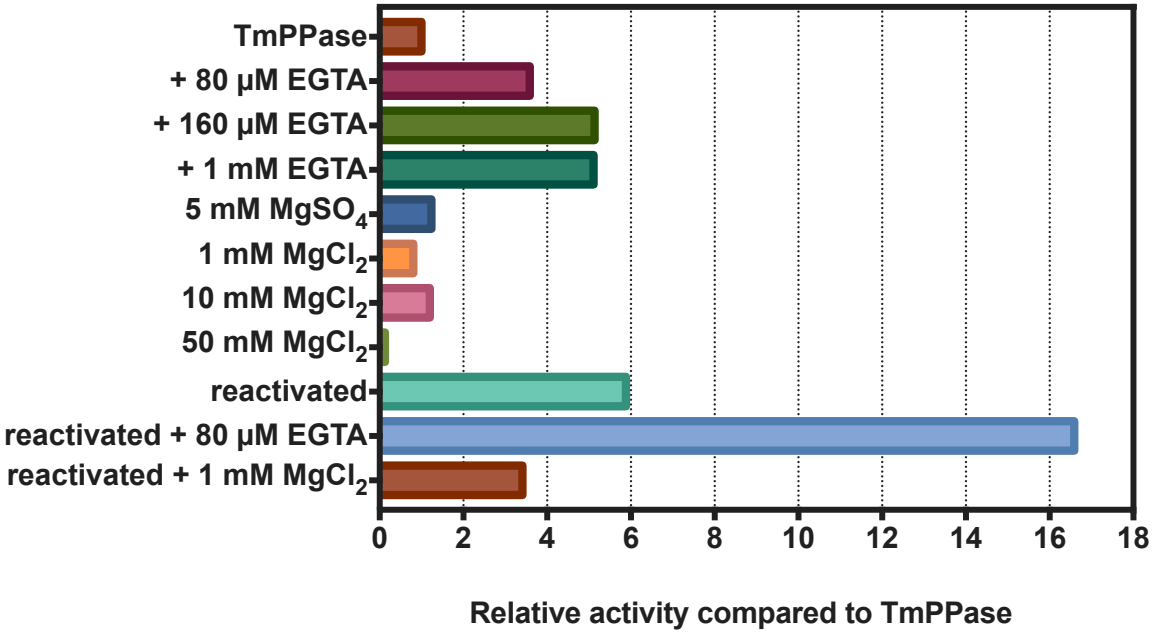


Figure 3.2: Hydrolytic Activity Assay of TmPPase Testing Calcium Inhibition.
Activity assay results for calcium inhibition experiments showing relative activity of samples compared with TmPPase sample with no supplements to the activity assay buffer. Chemical given in each lane represents either a substitution for $MgCl_2$, alterations to the concentrations of $MgCl_2$ or supplementation alongside $MgCl_2$ already present in activity assay buffer. Results are based on N=1 so cannot provide statistical significance.

3.3. Expression and Purification of PaPPase

3.3.1. Small-scale Expression and Purification Tests

3.3.1.1. Optimising Media Conditions for PaPPase Expression

The expression and purification of PaPPase was optimised to try and improve the yield and purity of the samples for crystallographic studies. Initially, four constructs were made, Pa A – Pa D (Table 2.1). Pa C and D contained a signal peptide from *Trypansoma cruzi*, which in previous studies had improved the yield of M-PPases expressed in yeast (Kellosalo *et al.*, 2011). Pa B and D contained a Glycine-Glycine (GG) linker between the His-tag and protein to try and improve column binding and elution during the purification stage. The Pa A construct, contained neither the signal peptide or a GG-linker and thus represented the construct used for previous expression experiments of PaPPase (Kellosalo *et al.*, 2011). Of these four planned constructs, only three were successfully generated (Pa A, Pa B and Pa C) and tested in expression trials. Due to the lack of difference in the resulting expression between each of the constructs, no further work was carried out in generating the fourth construct (Pa D).

Initial tests were carried out on the type of media used during expression, with optical density measurements at 600 nm (OD_{600}) used to measure the growth rate of PaPPase-expressing yeast cultures. There was no noticeable change in OD for the first 3-4 hours of expression post inoculation (Figure 3.3). In all expression samples the OD increased between 3 and 6 hours before reaching an apparent stationary phase at 6-7 hours at an OD_{600} of between 10 and 12. In cultures grown entirely in selective media (Figure 3.3a), the OD_{600} then decreased to a value between 6 and 8 until the point of harvesting, 8 hours post inoculation. The cultures switched from selective to rich media (Figure 3.3b) showed a sharp increase in OD_{600} from the 6 to 8 hour time points, post inoculation, at which time the OD_{600} had reached 14.

The secondary peak seen in the combined media method shows that after 6 hours the yeast culture enters a secondary exponential phase. This could be due to the loss of selective pressure, which allows the yeast cells to remove the plasmid or otherwise inactivate expression of the gene of interest. Expression of M-PPases was not beneficial to the growth of the yeast cells, either diverting resources away from vital processes or simply due to the build-up of ion-translocating proteins in the yeast cell membrane, which can disrupt native ion gradients required for survival. It was highly likely that the yeast growing between the 6 and 8-hour time-point were no longer expressing functional PaPPase. These results were from a single replication without repeat, so statistical significance cannot be determined for changes in expression. However, results of future experiments using only selective SCD-Leu media, provided sufficient selective pressure to maintain the PaPPase-containing plasmid within the yeast cells. There was no apparent difference in the growth curves of each individual PaPPase construct within the same culturing conditions. Pa C, containing the *T. cruzi* signal sequence, did not express any better than the two constructs lacking this sequence (Figure 3.3). Therefore, only the Pa A and Pa B constructs were taken forward for purification trials.

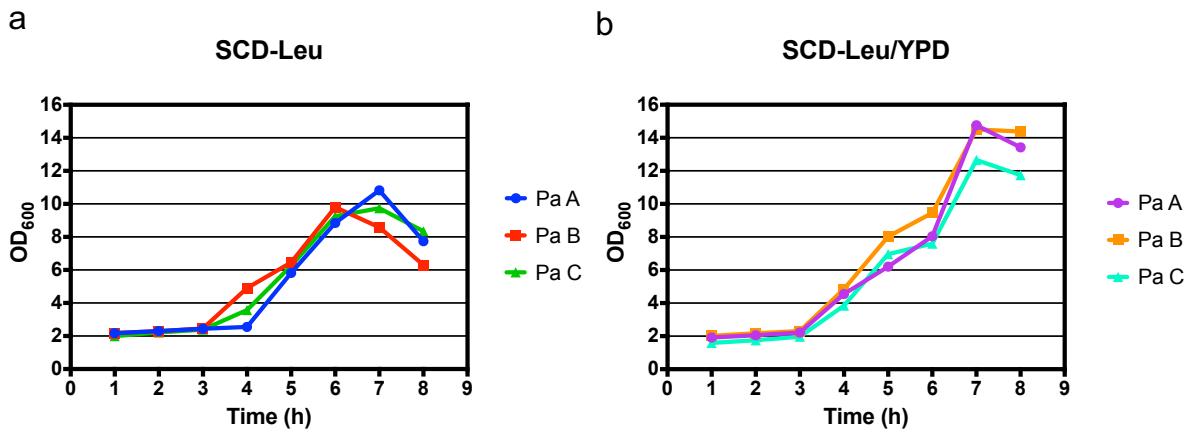


Figure 3.3: Growth Curves for PaPPase Construct Expression Test.
Growth curves of Pa A, Pa B and Pa C constructs involving overnight yeast cultures prepared in SCD-Leu media and transferred to (a) SCD-Leu media or (b) YPD media conditions for the 8-hour expression trial. Results are based on a single expression trial (N=1) so statistical significance cannot be determined.

3.3.1.2. Testing PaPPase Expression Constructs

Constructs Pa A and Pa B (without and with the GG linker) were expressed on a larger scale. Cell yields were roughly equal for both constructs (Table 3.1), but total membrane protein concentrations, measured by the Bradford assay, revealed a higher concentration in the Pa B sample over Pa A. The resulting protein was then subjected to an imidazole gradient to test the effect of the GG-linker during purification. Previous studies had shown that the PaPPase protein did not bind well to the column, leading to elution at very low concentrations of imidazole (~30 mM) (Kellosalo *et al.*, 2011). The GG-linker was designed to increase the distance between the His-tag and protein, leading to stronger column binding such that elution only occurs at higher concentrations of imidazole.

Table 3.1: Cell and Protein yields from PaPPase Construct Expression Test.
Yield information for Pa A and Pa B variants of PaPPase expressed in yeast and purified via Ni-NTA purification.

	Pa A	**Pa B**
Cell yield (g cells/L culture)	3.86	3.64
Total membrane protein concentration (mg/ml)	57.5	123.8
Protein yield (mg protein/L culture)	0.01	0.171

Analysis of the samples by SDS-PAGE showed a band at the expected molecular weight of PaPPase, around 65 kDa, for both the Pa A and Pa B samples (Figure 3.4). Comparing the purification of Pa A and Pa B, both samples start to elute in the 40 mM imidazole samples with the most prominent band in the 200 mM imidazole sample and with small

amounts present in the 300 mM imidazole eluted sample. Contaminating bands, present around the 25 kDa marker band, are still visible in the 30 mM imidazole sample but are not visible in the 40 mM or higher samples for both Pa A and Pa B. An additional band is present around the 245 kDa marker band and likely represents either an aggregated sample or a dimer of PaPPase. This band was also seen during previous expression and purification of PaPPase both in SDS-PAGE and Western blot experiments (Kellosalo, personal correspondence).

The final yield of protein obtained was 0.01 mg and 0.171 mg of Pa A and Pa B per litre of culture, respectively. Pa A had a lower overall yield than Pa B during solubilisation and purification. This could be due to issues encountered during yeast cell lysis, in which the Pa A sample was insufficiently cooled. This led to a significant decrease in active protein throughout the purification procedure. Following solubilisation any denatured protein would have been removed along with other aggregates during centrifugation steps and so other than the reduced yield, would not affect any other tests carried out on these protein samples, such as that of the imidazole gradient to test column binding.

Yields obtained from the purification of Pa A and Pa B were also quite low, 100-fold and 10-fold lower than TmPPase, respectively (Section 3.2). The yeast cultures were scaled up for expression, however the times used were from the small-scale expression trials. The cultures reached a much lower OD_{600} at the point of harvesting and thus less protein was expressed. This did not affect the testing of the constructs, but simply reduced the overall yield obtained. All subsequent cultures were grown for 10 hours, as opposed to 6 to ensure there was sufficient time for optimal expression of PaPPase.

Overall, the inclusion of the GG-linker did not improve the column-binding efficiency of these proteins during purification, however, after 2 washes with 20 mM imidazole there were no significant contaminating bands in either construct. This meant that future experiments utilised a two-step wash and elution method during purification, wherein the protein sample would be washed twice with buffer containing 20 mM imidazole and then eluted with 300 mM imidazole.

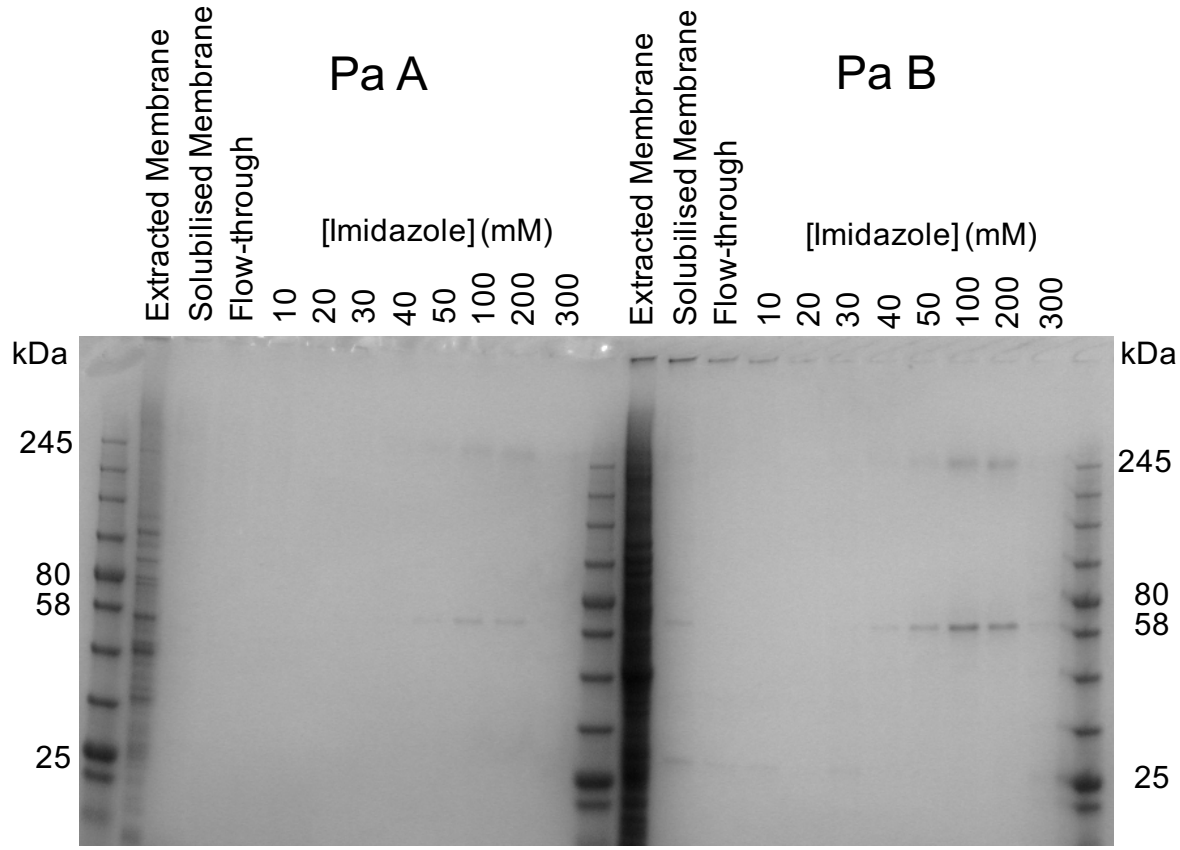


Figure 3.4: SDS-PAGE of PaPPase Elution Test.
Coomassie-stained SDS-PAGE showing elution test with an imidazole gradient from 20-300 mM imidazole during purification of Pa A and Pa B constructs.

3.3.1.3. Relipidation Trials for Optimising Hydrolytic Activity of PaPPase

Relipidation is a technique that has been used during the earlier purification studies on TmPPase and was required to replace lipids lost during purification, which were vital for the activity of the protein even in detergent micelles (López-Marqués *et al.*, 2005). TmPPase was supplemented with a final concentration of 12 mg/ml Soybean lecithin and 4.5 % DDM, leading to optimal activity following purification. Several relipidation methods were tested during a hydrolytic activity assay for each of the PaPPase variants, Pa A and Pa B. These included the method used above for TmPPase, a method used previously for the purification of PaPPase in which 0.036 mg/ml soybean lecithin was added (Kellosalo *et al.*, 2011), and an additional method in which 12 mg/ml soybean lecithin was added without extra detergent.

The specific activities of Pa A and Pa B without relipidation were 10 and 32 nmol P_i µg protein^{-1} min^{-1}, respectively (Figure 3.5). When supplemented with lipids and detergent, the activity increased to 11 and 42 nmol P_i µg protein^{-1} min^{-1}, respectively. However,

supplementation with only 47 μM lipids led to a decrease in activity to 3.4 and 5.4 nmol P_i μg protein^{-1} min^{-1}, respectively. The hydrolytic activity increased when the protein was supplemented with 12 mg/ml lipids, to 19.5 and 56.5 nmol P_i μg protein^{-1} min^{-1}, respectively. These results are from a single experiment without repeats and so the statistical significance of these results cannot be confirmed. It is still clear that the Pa B construct had a higher activity amongst all samples compared with Pa A and future experiments using Pa B supplemented with 12 mg/ml lipids showed a consistently high hydrolytic activity.

The Pa B variant consistently had a higher specific activity than Pa A and so was taken forward for subsequent experiments. It is possible this higher specific activity is due to the structural importance of the N-terminus on which the His-tag is located. It is possible that the His-tag prevents efficient folding of the N-terminus region. The addition of the GG-linker would then move the His-tag two residues further away and may allow for more efficient folding of the protein. This could improve the interactions made by the N-terminal residues to help stabilise the protein.

These relipidation results also show that the addition of extra detergent with the lipids actually decreases the overall activity of PaPPase compared with using lipids alone (Figure 3.5). If these lipids are required for activity then the addition of extra detergent may prevent the lipids from binding to the protein, leading to a lower activity. Based on the activity assay results, future expression and purification experiments were carried out using only the Pa B construct. Moreover, PaPPase was supplemented with 12 mg/ml soybean lecithin, following purification, to recover the activity lost by the removal of lipids during purification.

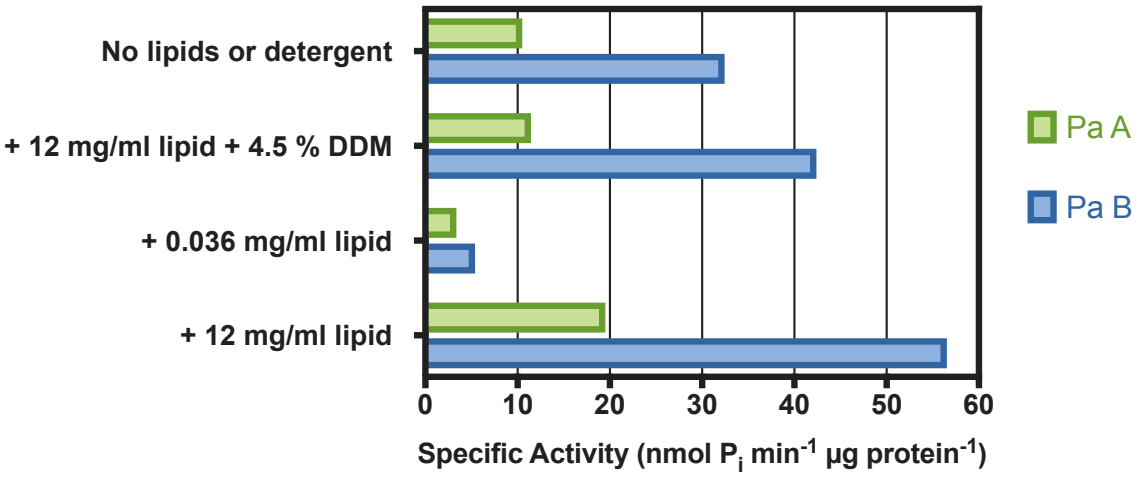


Figure 3.5: Hydrolytic Activity Assay testing Relipidation of PaPPase.
Activity assays results showing specific activity of Pa A and Pa B protein samples and comparisons with different post-purification relipidation methods. Results are from a single experiment without repeats (N=1).

3.4. Optimisation of Solubilisation Conditions for PaPPase

3.4.1. Pa B Large Scale Expression

3.4.1.1. Expression of PaPPase in Yeast

Large-scale expression of PaPPase was carried out in 10 L batches. OD_{600} was monitored for the cultures over the course of 10 hours (Figure 3.6). This showed a linear increase in OD over the time course until the maximal value of 10 is reached at the 10-hour time-point. The cells were harvested and the total yield was 5.84 g cells/L culture. Following membrane extraction, the total protein concentration of the extracted membrane sample was determined to be 126.6 mg/ml using the Bradford assay.

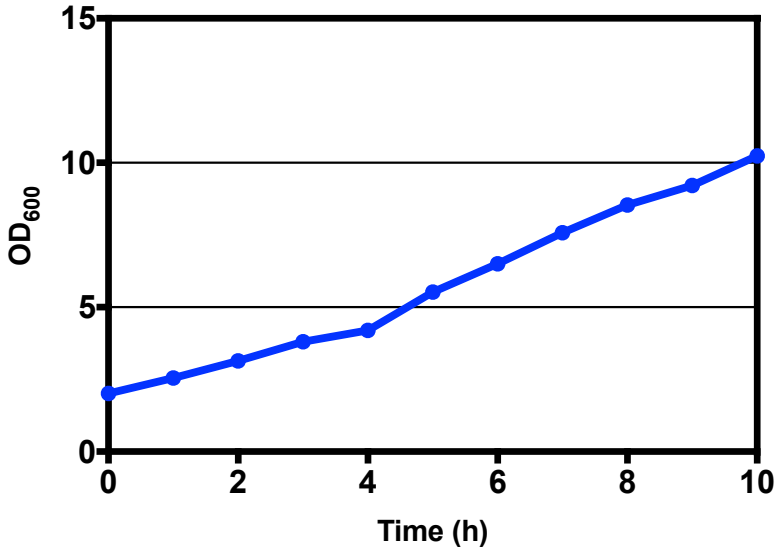


Figure 3.6: Growth Curve for Large-scale PaPPase Expression.
Growth curve showing OD_{600} for expression of Pa B construct (PaPPase) in 2 L shake-flasks over a 10 hour time course. Results are from a single experiment without repeats (N=1) so no statistical significance can be determined.

3.4.1.2. Detergent Screen for Solubilisation of PaPPase

Solubilisation of PaPPase had previously been carried out in DDM (Kellosalo *et al.*, 2011) but only a handful of other detergents had been tested during crystallisation trials (Kellosalo *et al.*, 2013). In this study, a screen of 20 different detergents was tested on PaPPase to see if these would yield a higher proportion of solubilised protein compared to DDM (Table 3.2). From this screen, solubilisation was observed via Western blot for β-D-Fructopyranosyl-α-D-Glucopyranoside Monododecanoate (Sucrose monododeconoate), Tetraethylene Glycol Monooctyl Ether (C8E4), Hexaethylene Glycol Monooctyl Ether (C8E6), Octaethylene Glycol Monododecyl Ether (C12E8), n-Dodecyl−β−D-Maltopyranoside (DDM), n-Undecyl−β−D-Maltopyranoside (UM), n-Tetradecyl−β−D-Maltopyranoside (tetraDM), n-Decyl−β−D-Thiomaltopyranoside (DTM), n-Dodecyl−β−D-Thiomaltopyranoside (DDTM), n-Nonyl−β−D-Thiomaltopyranoside (NTM), n-Octyl−β−D-Thiomaltopyranoside (OTM), n-Undecyl−β−D-Thiomaltopyranoside (UTM) and n-Nonyl−β−D-Glucopyranoside (NG) (Figure 3.7). UDM and DTM provided the most efficient solubilisation (~80 %), with NTM, NG, OTM, tetraDM, C8E4 and C12E8 showing ~50 % solubilisation. A low degree of solubilisation (< 10 %) was observed for 3-Cyclohexyl-1-Propyl-β-D-Glucoside (CYGLU-3), Methyl-6-O-(N-Heptylcarbamoyl)-α-D-Glucopyranoside (Anameg-7), n-Heptyl−β−D-Thioglucopyranoside (HTG), 7-Cyclohexyl-1-Heptyl-β-D-

Maltoside (Cymal-7), n-Tridecyl−β−D-Maltopyranoside (TriDM) or 2,6-Dimethyl-4-Heptyl−β−D-Maltopyranoside. The remaining detergents showed minor solubilisation (10-40 %) (Table 3.2; Figure 3.7).

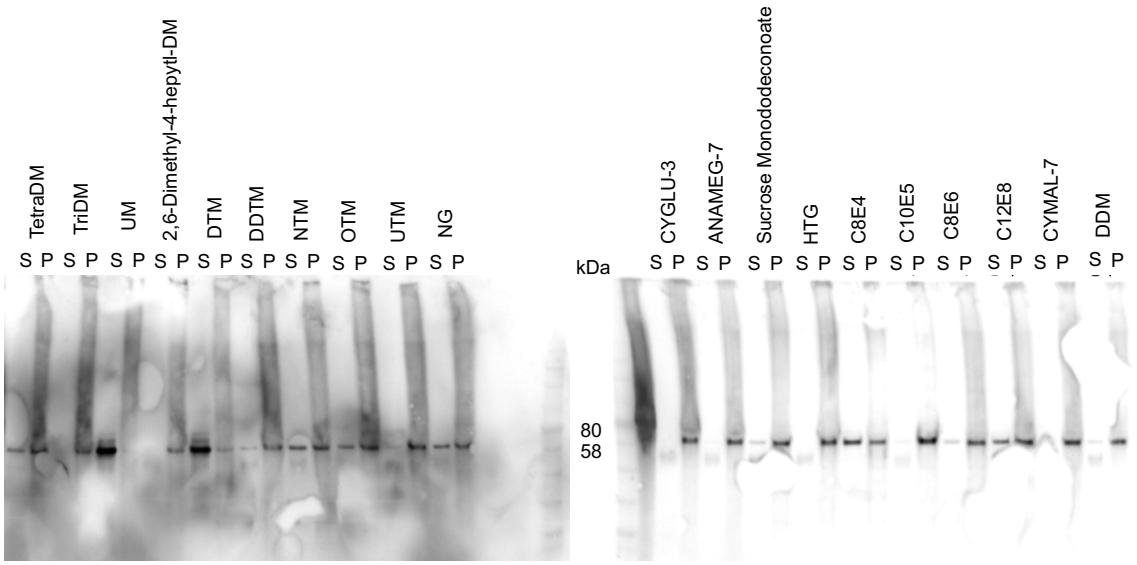


Figure 3.7: Western Blot for Detergent Solubilisation Screen of PaPPase.
Western blot analysis of small-scale detergent solubilisation test using anti-mPPase antibody showing both the supernatant and pellet fractions following solubilisation at 70 ˚C.

Table 3.2: Detergent Solubilisation Screen of PaPPase.

List of detergents used in the PaPPase solubilisation screen along with abbreviated names, degree of solubilisation as determined via Western blot analysis and relative hydrolytic activity.

Name of detergent	**Abbreviated name**	**Solubilisation (%)**	**Relative activity compared to DDM (%)**
7-Cyclohexyl-1-Heptyl-β-D-Maltoside	CYMAL-7	< 10	109.0
n-Octyl−β−D-Thiomaltopyranoside	OTM	~ 50	94.8
n-Nonyl−β−D-Thiomaltopyranoside	NTM	~ 50	102.5
n-Decyl−β−D-Thiomaltopyranoside	DTM	~ 80	102.6
n-Undecyl−β−D-Maltopyranoside	UM	~ 80	109.1
n-Undecyl−β−D-Thiomaltopyranoside	UTM	10 – 40	100.1
n-Dodecyl−β−D-Maltopyranoside	DDM	10 – 40	100.0
n-Dodecyl−β−D-Thiomaltopyranoside	DDTM	10 – 40	99.9
n-Tridecyl−β−D-Maltopyranoside	TriDM	< 10	95.7
n-Tetradecyl−β−D-Maltopyranoside	TetraDM	~ 50	96.8
2,6-Dimethyl-4-Heptyl−β−D-Maltopyranoside	2,6-Dimethyl-4-Heptyl−β−D-M	< 10	111.5
3-Cyclohexyl-1-Propyl-β-D-Glucoside	CYGLU-3	< 10	109.0
n-Nonyl−β−D-Glucopyranoside	NG	~ 50	94.6
Methyl-6-O-(N-Heptylcarbamoyl)-α-D-Glucopyranoside	Anameg-7	< 10	105.2
β-D-Fructopyranosyl-α-D-Glucopyranoside Monododecanoate	Sucrose monododeconoate	10 – 40	110.6
n-Heptyl−β−D-Thioglucopyranoside	HTG	< 10	100.6
Tetraethylene Glycol Monooctyl Ether	C8E4	~ 50	104.1
Pentaethylene Glycol Monodecyl Ether	C10E5	10 – 40	103.3
Hexaethylene Glycol Monooctyl Ether	C8E6	10 - 40	25.1
Octaethylene Glycol Monododecyl Ether	C12E8	~ 50	111.0

All detergents, with the exception of C8E6, showed hydrolytic activity (Figure 3.8). UDM, 2,6-Dimethyl-4-heptyl-B-DM, CYGLU-3, Anameg-7, sucrose monododeconoate, C12E8 and CYMAL-7 have activity above the DDM-solubilised sample (105-110 % activity). TetraDM, TriDM, OTM and NG have activities below that of DDM (~90 %), with C8E6 having around 25 % activity compared to the DDM sample. However, there are limitations to the hydrolytic activity assay carried out on solubilised membrane samples, since phosphate is used to stabilise the protein during solubilisation at 70 ºC and this can carry over into the activity assay and produce background signal that can mask the results. In addition, activity assay results were from a single experiment without repeats, meaning that statistical significance cannot be determined. However, all samples, with the exception of C8E6 had clearly similar activities and so no major conclusions could be drawn from this experiment regardless.

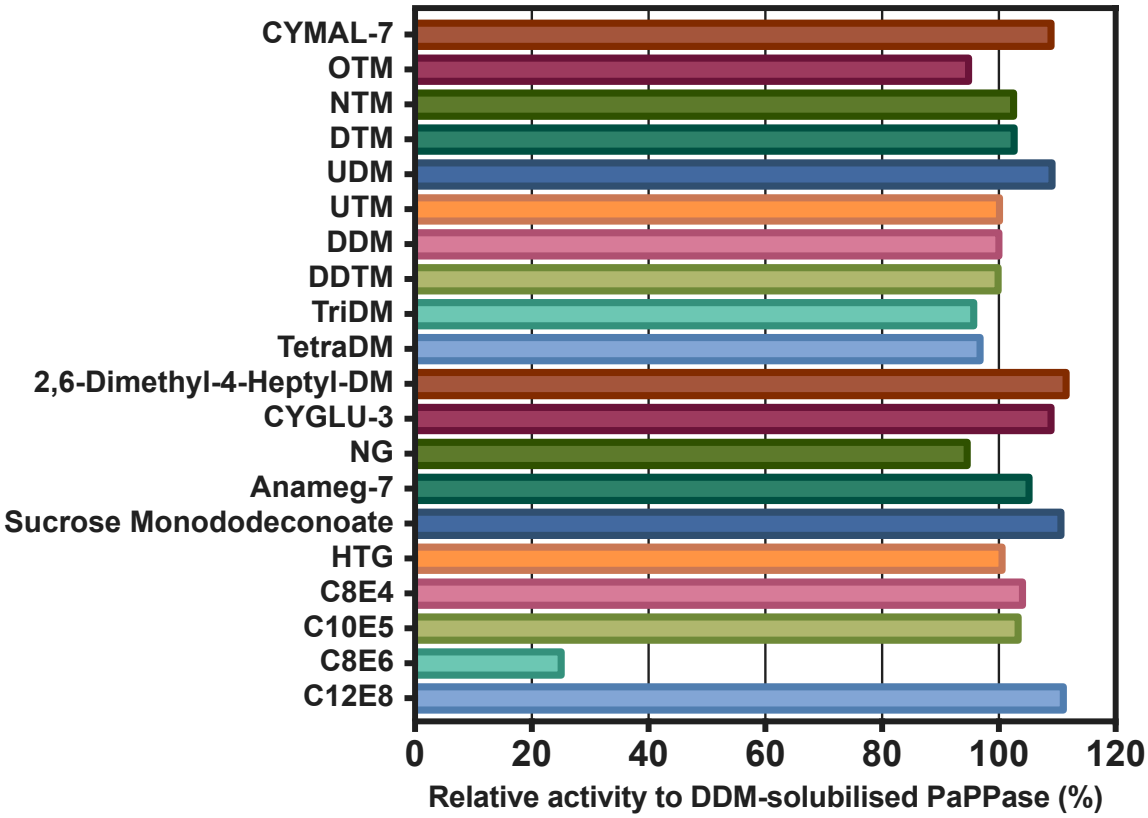


Figure 3.8: Hydrolytic Activity Assay for Detergent Solubilisation Screen of PaPPase.
Activity assay results for detergent solubilisation screen of PaPPase samples showing relative activity compared with DDM-solubilised PaPPase protein. Results are from a single experiment without replication (N=1) so statistical significance cannot be determined.

3.4.1.3. Larger-scale Solubilisation Tests of PaPPase

Two of the best detergents from the small-scale solubilisation screen were chosen for scaling up, alongside DDM as a control. These were DM and UM and were chosen due to the high degree of solubilisation observed during the small-scale solubilisation test (70-80 % solubilisation). Alongside testing the detergents, PP_i and IDP were also used to try and test their effect on stabilising the protein during solubilisation at 70 ºC. This was carried out to try and mitigate the effect of phosphate crossing over into the activity assay carried out on solubilised protein samples.

The total activity of each solubilisation sample showed that all samples solubilised in the presence of PP_i had the highest activity (Figure 3.9). Samples solubilised without any phosphate compounds showed the lowest activity and those solubilised in the presence of

IDP had activity between the PP_i and water samples. Moreover, the samples solubilised in DM and UM both showed higher activity following solubilisation than those solubilised in DDM. This assay looked at total activity of the protein (nmol P_i . min^{-1}) and does not factor in protein concentration as this is difficult to accurately measure at this stage of the purification. This affected the comparisons made between each of the detergent-solubilised samples but did not affect any conclusions drawn between samples within the same detergent solubilisation group, as these all used identical samples for testing.

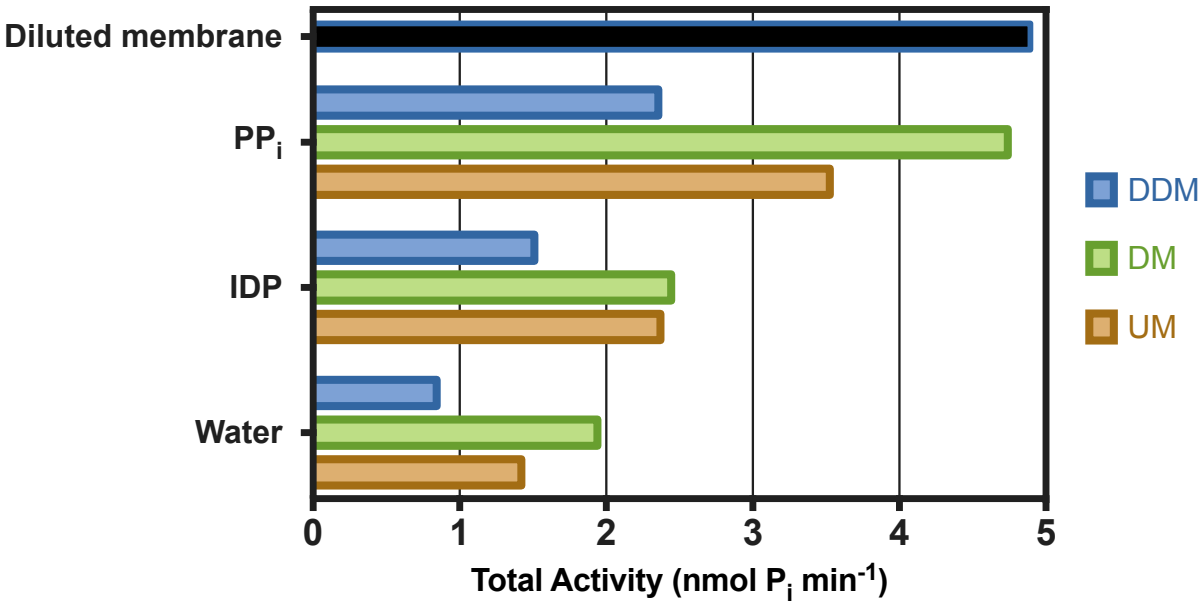


Figure 3.9: Hydrolytic Activity Assay for Large-scale Solubilisation Test of PaPPase.
Activity assay results for larger-scale solubilisation test of PaPPase using DDM, DM and UM alongside different phosphate-containing compounds to compare total activity between each sample. Results are from a single experiment without replication (N=1).

The addition of a phosphate-containing compound was definitely required to keep the protein stable during solubilisation at 70 ºC. The increased activity seen in the PP_i samples may also be due to cross-over into the activity assay buffer. PaPPase will hydrolyse PP_i during the solubilisation step as it has an optimal activity temperature of around 80 ˚C. Therefore, PP_i or P_i present in the solubilisation buffer will also be present in the activity assay samples. Since the activity assay measures P_i concentrations in the buffer, this artificially inflated the activity results over the samples tested with IDP or water. Moreover, the addition of IDP during solubilisation also carried over and would have had the opposite effect by inhibiting the activity of the protein. Therefore, the rise in activity

seen with the IDP sample over the no-phosphate control (Figure 3.9) shows this increase was caused by a genuine stabilising effect of the phosphate-compound during solubilisation at 70 ˚C.

Although this experiment was inconclusive on whether PP_i or IDP was best for solubilisation, it highlighted the importance of ensuring that at least one of these compounds was added as activity was definitely lost during solubilisation in their absence. The results of this experiment mean that future solubilisation experiments were carried out using DM and using PP_i to stabilise the protein, since IDP would also have a definite inhibitory effect on the activity as opposed to PP_i. If activity assays are required on solubilised protein samples, the crossover of phosphate into these samples must be factored into the results obtained, as its addition is unavoidable.

3.5. Optimising Purification of PaPPase

3.5.1. Large-scale Expression and Purification of PaPPase in Yeast

The final stages of purification optimisation involved exchanging PaPPase into various detergents in order to carry out multiple crystallisation trials. The detergent can be exchanged on the Nickel column by washing and eluting the protein in buffer containing the detergent of interest. A range of detergents had previously been tested for PaPPase in terms of hydrolytic activity and crystallisation, with the best combination found to be DM, producing crystals that diffracted to 5.8 Å (Kellosalo *et al.*, 2013; Kellosalo *et al.*, 2011). A new set of detergents was selected to carry out additional studies, including DDM, DM, OGNG, LMNG, OM, NM and NG. These detergents were selected as they contained different overall structures, such as LMNG and OGNG or contained shorter acyl chain lengths that could potentially improve the crystallisation of this protein. Cell yields from large-scale expression of PaPPase were between 4 and 6.6 g cells/L yeast culture (average 5.04 g/L). Samples were combined and membrane extractions were carried out with the resulting total membrane protein concentration found to be 161.1 mg/ml using the Bradford assay.

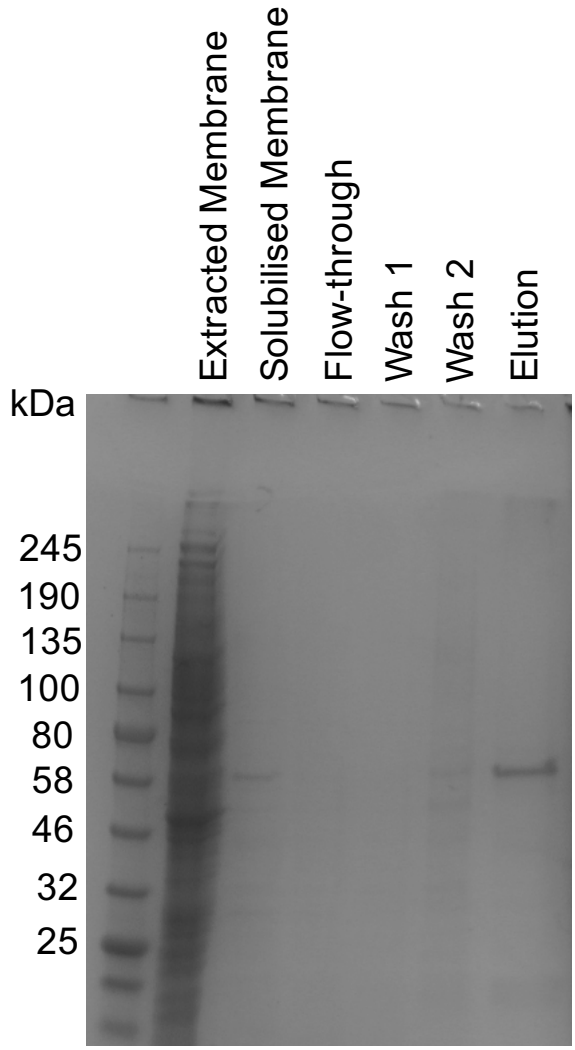


Figure 3.10: SDS-PAGE of PaPPase Purification.
Coomassie-stained SDS-PAGE gel of a large-scale purifications of PaPPase following two-step imidazole wash and elution.

This sample was used for multiple purifications, being exchanged into the required detergents. SDS-PAGE of the DM-purified sample (Figure 3.10) showed the protein sample was pure, with only a single band at the expected molecular weight, between the 58 and 80 kDa marker bands. PaPPase samples were also purified in a range of detergents, including OGNG, LMNG, DDM and DM (Figure 3.11), which had specific activities of 90, 132, 98 and 40 nmol PO_4 μg protein^{-1} min^{-1}, respectively. In addition, PaPPase was also exchanged into OM, NM and NG in preparation for crystallisation trials, with SDS-PAGE showing a band at the expected molecular weight of PaPPase and another around the 245 kDa marker band, representing either aggregated or dimerised PaPPase (Kellosalo, Personal correspondence) (Figure 3.12). As with TmPPase, the final yields of PaPPase in various detergents were lower than seen in the previous study (Kellosalo

et al., 2011), in which the maximal yield obtained was around 1 mg protein/L culture. The yields obtained in this study were between 0.4 and 0.5 mg protein/ L culture, within acceptable levels given the differences in estimating protein concentration using the A280 method (Section 3.2).

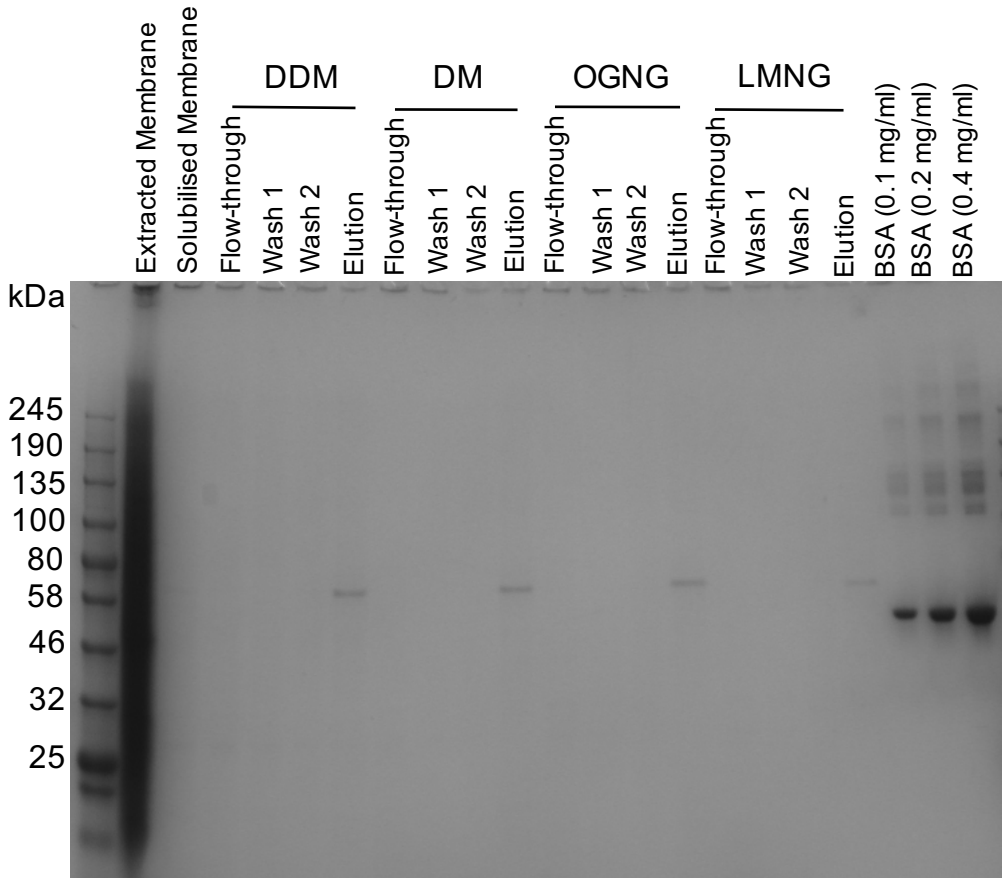


Figure 3.11: SDS-PAGE of Initial Detergent Exchange During PaPPase Purification.
Coomassie-stained SDS-PAGE showing results of on-column detergent exchange of PaPPase into DM, OGNG and LMNG during purification.

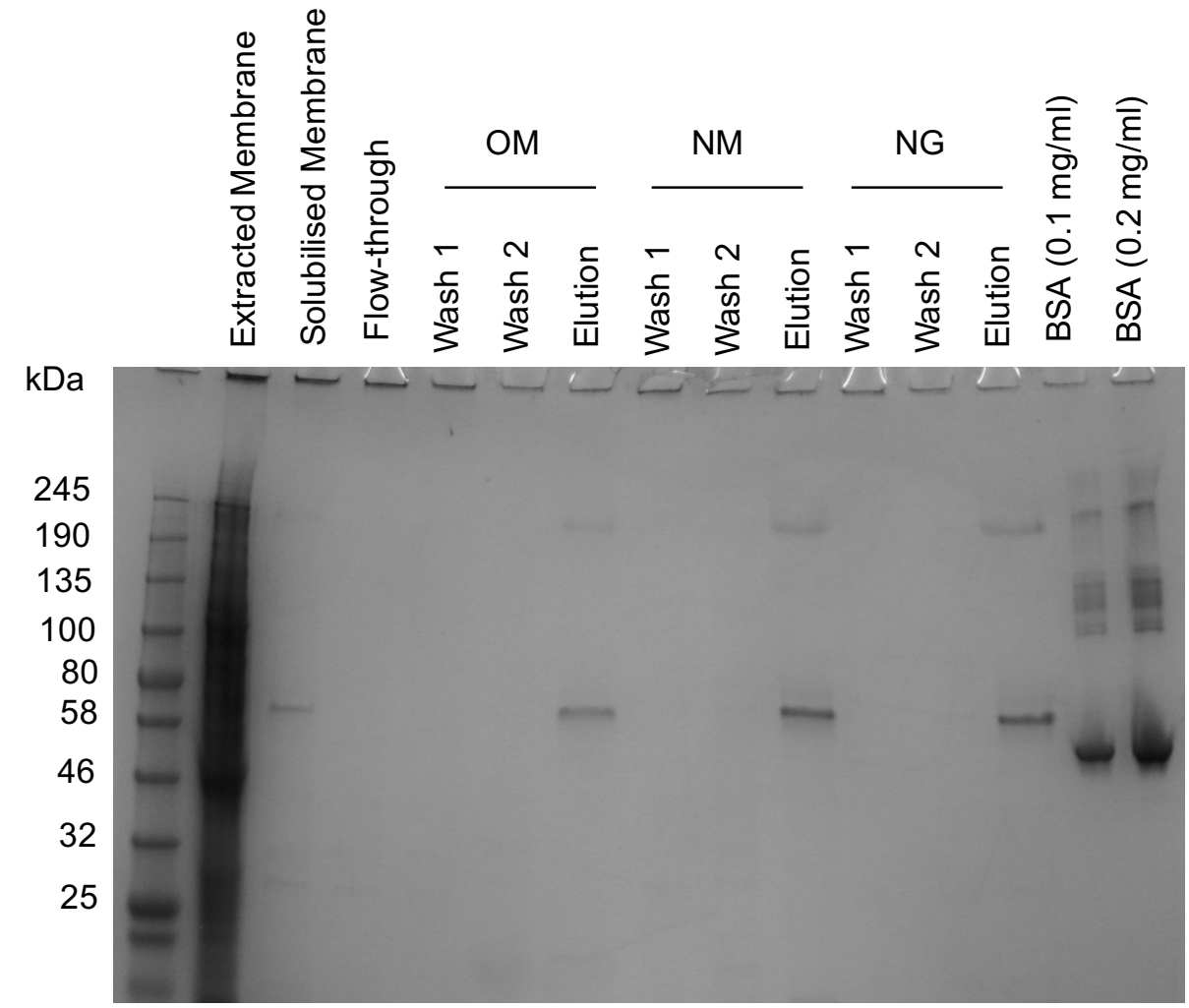


Figure 3.12: SDS-PAGE of Additional Detergent Exchange During PaPPase Purification.
Coomassie-stained SDS-PAGE gel showing results of on-column detergent exchange of PaPPase into OM, NM and NG during purification.

3.6. Mutagenesis Experiments on PaPPase

3.6.1. C-terminal Truncation Mutant

A sequence alignment of PaPPase with other M-PPases had revealed the C-terminus of PaPPase was longer than both TmPPase and VrPPase, containing 13 extra residues. In TmPPase, these residues form stabilising interactions with the end of helix 14 and are all clearly visible in the structure (Figure 3.13) (PDB ID: 4AV3). These residues are also close the dimer interface and so may also play a role in the stabilisation of the dimer in TmPPase.

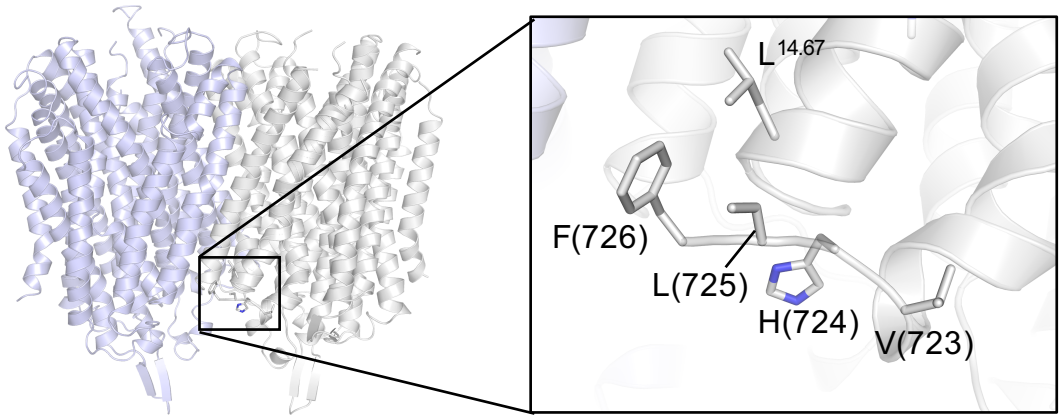


Figure 3.13: Overview of C-terminus in TmPPase.
Dimer representation of TmPPase:CaMg (PDB ID: 4AV3) showing the location of the C-terminus in chain A (grey) and an enhanced view of the C-terminus showing the final four residues of helix 16 and their interactions with $L^{14.67}$. Loop residues are labelled with the TmPPase numbering system.

These extra residues in PaPPase may reduce the efficiency of crystal packing if they form no interactions with other parts of the protein. These residues were mutated out of PaPPase to see if this would improve the crystallisation and resulting diffraction of crystals compared with the wild-type. In-Fusion cloning was used to remove the 13 residues from the C-terminus of PaPPase and sequencing was used to confirm the presence of the PaPPase gene without the C-terminal residues. After cloning, the protein was expressed and purified alongside the wild-type PaPPase and SDS-PAGE showed a band for this sample around the expected molecular weight, slightly lower than the wild-type protein (Figure 3.14). Additional bands were also visible in the gel at lower molecular weights than the expected PaPPase mutant. These may be degradation products of PaPPase as these mutants may not be as thermostable as the wild-type protein. Since structures were not being sought of these mutants, functional assays were carried out on the mutant protein samples. The protein yield was also calculated via the A280 method and found to be 0.57 mg/L culture. The specific activity of the wild-type and C-terminal truncated mutant of PaPPase was found to be 447.2 and 39 nmol P_i µg $protein^{-1}$ min^{-1}, respectively (Figure 3.15), a significant decrease in hydrolytic activity upon removing the C-terminal residues (*P*-value = 0.0067).

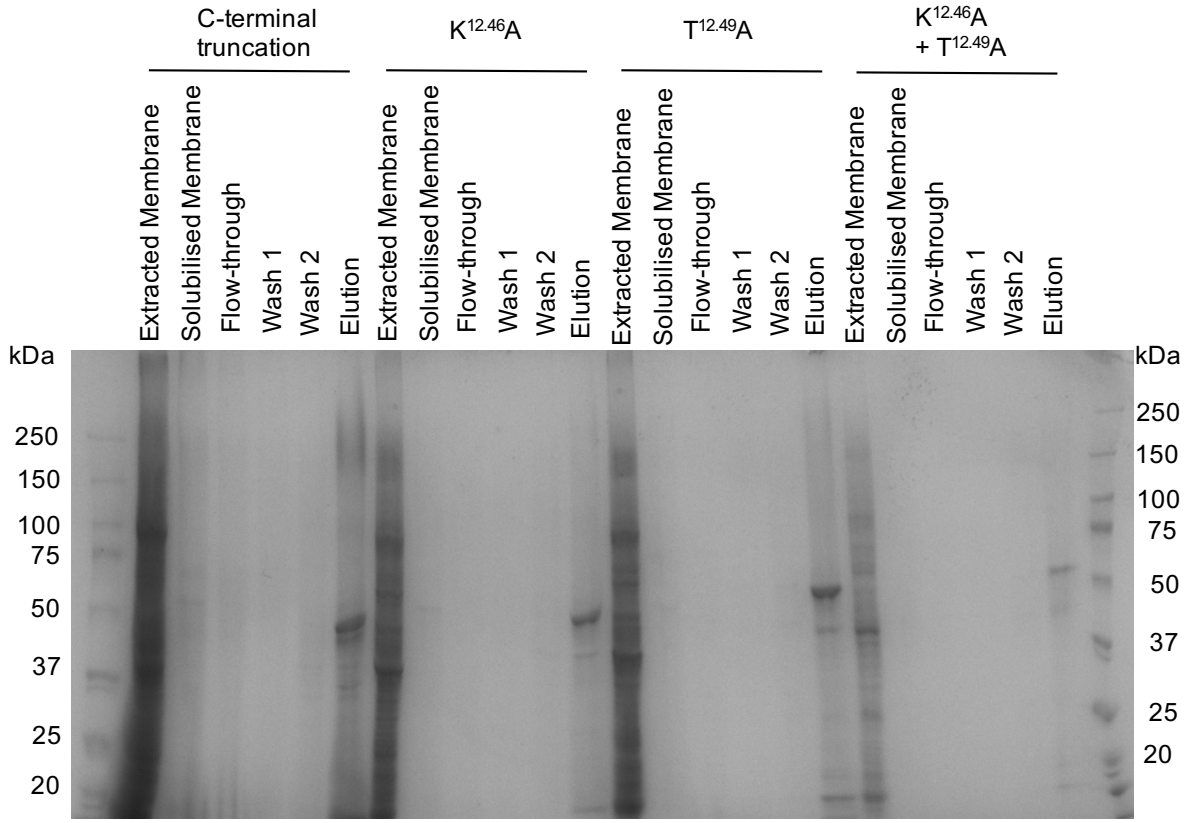


Figure 3.14: SDS-PAGE of PaPPase mutant Purification.
Coomassie-stained SDS-PAGE showing purification of PaPPase mutants, including C-terminal truncation mutant and K^+-dependent mutants.

The decrease in hydrolytic activity may show that the 13 C-terminal residues actually play an important role in the structure of the protein, stabilising it by forming interactions with other residues. This was observed in the structure of TmPPase (Kellosalo *et al.*, 2012), where it was shown that addition of purification tags to the C-terminal end of TmPPase disrupted folding and so produced unfolded and inactive protein (Kellosalo, personal communication). These residues may also be important for the thermostability of PaPPase, with the mutant becoming more susceptible to denaturation during solubilisation at 70 ˚C. Since the protein is not active and potentially unfolded, crystallisation experiments were not carried out. However, future experiments could look into different lengths of C-terminal truncations to specifically see which residues are essential for stability of the protein and to see if any of these would help improve the crystal packing during crystallisation.

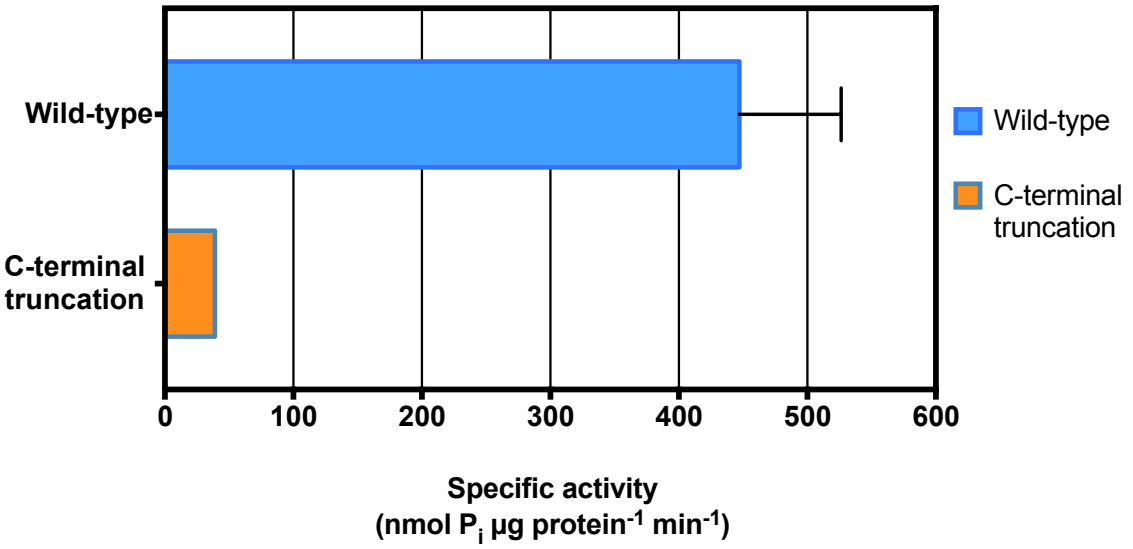


Figure 3.15: Hydrolytic Activity Assay of PaPPase C-terminal Truncation Mutant.
Specific hydrolytic activity results for PaPPase wild-type against the C-terminal truncation mutant, following relipidation. Experiments were carried out in triplicate, without repeats. Error bars denote standard error and are present for both the wild-type and C-terminal truncation samples.

3.6.2. K^+-Dependent Mutants

The study of potassium dependence in M-PPases is important to aid our understanding of how these proteins function. Several plant and parasite species include both potassium dependent and independent variants. It was speculated that the presence of both was because of the need to generate proton gradients over a range of different environmental conditions (McIntosh *et al.*, 2001). A previous study had generated several mutants of a potassium-dependent H^+-PPase, ultimately generating a potassium-independent variant (Belogurov & Lahti, 2002). This involved the mutations of $A^{12.46}$ to a lysine. It was found that a further mutation of $A^{12.49}$ to a threonine could modulate the affinity for potassium but did not itself generate a potassium-independent mutant.

The reverse of the potassium independence mutations, $K^{12.46}A$, $T^{12.49}A$ and a dual mutant, were generated in PaPPase and the presence of the mutations was confirmed via sequencing. Proteins were expressed using the standard expression and purification protocols for PaPPase (Section 3.5). The resulting yields of the $K^{12.46}A$, $T^{12.49}A$ and dual mutant were 1.32, 1.55 and 0.79 mg/L culture, respectively. SDS-PAGE of the mutants showed purified protein at the expected molecular weight of PaPPase (Figure 3.14).

The hydrolytic activity was measured both in the presence and absence of potassium alongside the wild-type for comparison. The specific activity for the wild-type PaPPase was 447 ± 79 and 403 ± 38 nmol P_i μg protein^{-1} min^{-1} with and without potassium, respectively, showing no drop in activity in the absence of potassium (Figure 3.16). The same was found for the $K^{12.46}A$ mutant, with activities of 437 ± 102 and 455 ± 153 nmol P_i μg protein^{-1} min^{-1} in the presence and absence of potassium, respectively (Figure 3.16). This differs from previous work which found that the A to K mutation was sufficient to generate a potassium-independent M-PPase. The activity of the $T^{12.49}A$ mutant in the presence of potassium is 300 ± 12 nmol P_i μg protein^{-1} min^{-1} and this drops to 203 ± 27 nmol P_i μg protein^{-1} min^{-1} in the absence of potassium (Figure 3.16). This also shows no change in the potassium dependence and was consistent with the previous findings.

The dual mutant had an activity of 205 ± 23 nmol P_i μg protein^{-1} min^{-1} in the presence of potassium (a statistically significant decrease compared to the wild-type with potassium, *P*-value = 0.04) and dropped significantly (*P*-value = 0.008) to 14 ± 14 nmol P_i μg protein^{-1} min^{-1} in the absence of potassium. The dual-mutant was only active in the presence of potassium, which showed that this mutant had been changed from a potassium-independent to a potassium-dependent M-PPase. However, the maximal activity achieved was still less than the wild-type protein. It is unlikely the overall structure of the dual-mutant has been completely destroyed as the activity is only lost in the absence of potassium, but this implied that other residues are involved in the potassium-independent mechanism of PaPPase.

The differences seen between the current and previous study (Belogurov & Lahti, 2002) could be for several reasons. Firstly, these two studies were carried out on the M-PPases from two distinct organisms, one bacterial and one archaeal. This means that the mutations made could alter any number of factors within these proteins, including affinity for substrate or even in some way uncoupling the hydrolytic and pumping functions, since the activity assay only measured hydrolytic activity. Secondly, these mutations might also affect the biochemical properties and not just the structure, for example, altering the hydrogen bonding patterns of the central channel, which may affect the activity of the protein. These results showed interesting preliminary findings that that both $K^{12.46}$ and $T^{12.49}$ are important for the potassium-independence of PaPPase. However, comparisons

of structural information with the wild-type structure would help to further validate these findings. In addition, a range of other mutations could also be tested on residues around this region of the protein, or mutations of $K^{12.46}/T^{12.49}$ to other residues with different biochemical properties such as non-polar or negatively charged residues, to see if these have other effects on the mechanism of potassium-independence.

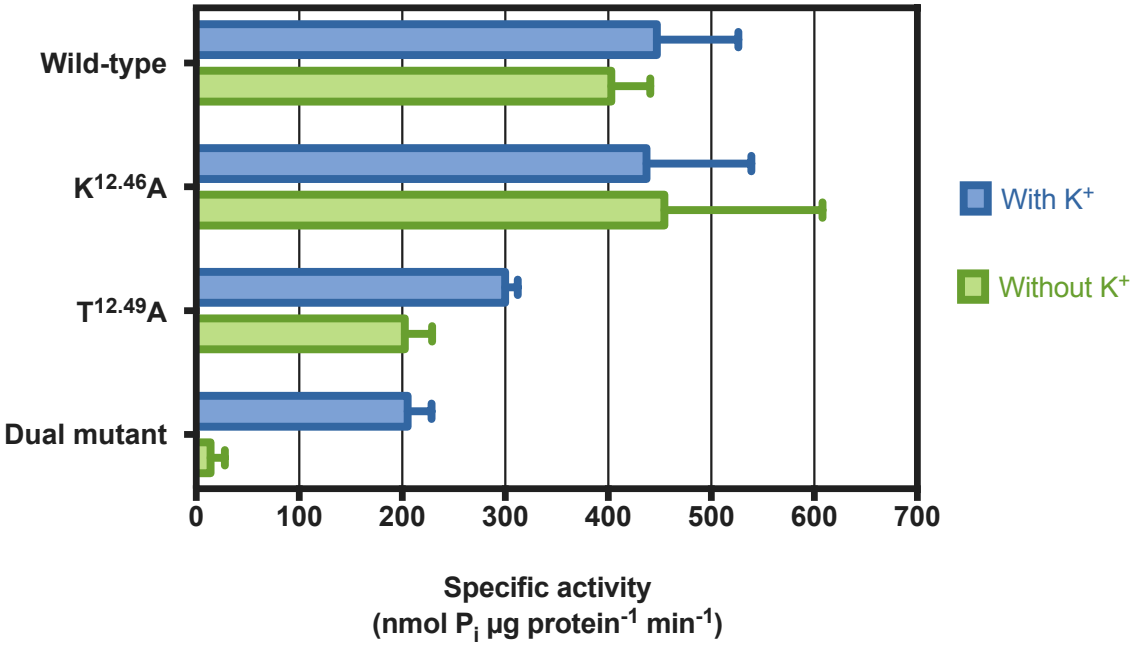


Figure 3.16: Hydrolytic Activity Assay of Potential K^+-dependent Mutants of PaPPase.
Specific hydrolytic activity results for the three potential K^+-dependent mutants of PaPPase compared with the wild-type PaPPase. All experiments were carried out in triplicate without repeats and error bars denote standard error from the mean.

3.7. Summary of M-PPase Expression and Purification

The aim of this chapter was to express and purify TmPPase as well as optimise the expression and purification of PaPPase. This was carried out to ensure sufficient pure protein was available for downstream structural and functional characterisation of these proteins. In addition, several mutational studies were carried out to test various biochemical effects, particularly looking at the potassium independence mechanism of PaPPase, as well as testing the removal of residues on the structural stability of this protein. Overall, in this chapter I showed expression of both TmPPase and PaPPase in yeast, solubilisation of the proteins in various detergents and purification of them to remove any contaminating proteins. In particular, the expression and purification of PaPPase was optimised to produce high concentrations of pure protein that were suitable

for crystallisation trials (see chapter 4). The removal of 13 C-terminal residues had a significantly negative effect on the activity of PaPPase and mutations in both $K^{12.46}A$ and $T^{12.49}A$ also produced a potassium-dependent variant of PaPPase.

Ultimately, both the TmPPase and PaPPase proteins were expressed and purified at concentrations suitable for continued crystallographic and biochemical characterisation. TmPPase was purified between 0.9 and 1.2 mg/L culture, providing sufficient protein to proceed to the Tryptophan fluorescence studies. PaPPase was found to remain pure and active when purified in a range of different detergents, including DDM, DM, OGNG, LMNG, OM, NM and NG, and was purified at concentrations between 0.4 and 0.5 mg/L culture, providing sufficient protein for crystallisation trials and tryptophan fluorescence studies. Issues of hydrolytic activity due to calcium inhibition were overcome by the addition of EGTA, which will improve the accuracy and reliability of future purification and activity assays on M-PPases. In addition, mutated variants of PaPPase ($K^{12.46}A$ and $T^{12.49}A$) highlighted important residues involved with the potassium-independent mechanism of PaPPase, providing additional information to help with future structural analysis. Overall, there was sufficient yields of pure protein to proceed on to structural studies of TmPPase and PaPPase using X-ray crystallography.

Chapter 4:

Structural Studies of Membrane Pyrophosphatases

4.1. Introduction to M-PPase Structutral Studies

The aim of the work in this chapter was to solve the structures of both TmPPase and PaPPase in as many different catalytic states as possible, such that I could make detailed comparisons and generate a model of the catalytic cycle. To this end, TmPPase had been expressed, purified and crystallised previous by Juho Kellosalo (Helsinki) and data collected was used in the current study to solve the TmPPase:Mg_5IDP and TmPPase:$MgWO_4$ structure. PaPPase was expressed and purified in yeast as part of the work in Chapter 3 and was used to carry out structural studies. There were no previous structures of PaPPase or any K^+-independent H^+-PPases, and therefore any structural information obtained would provide valuable insight into the mechanism of K^+-dependence in M-PPases. These structures could provide additional evidence to explain previous mutational studies, wherein a K^+-independent M-PPases could be generated via a single alanine-lysine mutation at position A/$K^{12.46}$ (Belogurov & Lahti, 2002).

The overall architecture of M-PPases was known using information from three X-ray crystal structures, including two of TmPPase, in the resting state and two-product bound state (Kellosalo *et al.*, 2012), and one of the M-PPase from *Vigna radiata* (VrPPase) in the substrate-analogue-bound state with IDP in the active site (Lin *et al.*, 2012). Despite this information, the specific mechanisms associated with each state of the catalytic cycle are still not completely understood. New structures of these proteins in other catalytic states will help us to understand the functional mechanisms of M-PPases and allow us to piece together a complete picture of the catalytic cycle.

Protein used in the generation of the TmPPase structures was already expressed, purified and crystallised by Juho Kellosalo (Helsinki). Two datasets had been collected and I processed these to obtain the final structures discussed in this chapter. The first was of TmPPase, to which the inhibitor and substrate analogue imidodiphosphate (IDP) had been added to the crystallisation conditions. The other dataset had been used for phasing of higher resolution structures during a previous study (Kellosalo *et al.*, 2012) but had never been refined or solved.

The aim of this chapter was to try and solve the two TmPPase datasets to obtain new information that could assist in the generation of the catalytic cycle model. Using this

information to see if these structures correlated with the previously solved structures and see if any new information could be obtained from these datasets. Moreover, it was hoped that the structure of PaPPase could be solved using X-ray crystallography, at a resolution that allows for comparison of both helical and residue changes between catalytic states and the different types of M-PPases.

4.2. Structural Studies of TmPPase Using X-ray Crystallography

4.2.1. Data Collection and Structural Refinement

Two new structures were solved of TmPPase, one in complex with the inhibitor and substrate analogue (Mg_5IDP) and the other in complex with the single-product-analogue ($MgWO_4$). The data were collected by a former graduate student (Dr. Juho Kellosalo, University of Helsinki) and data processing and refinement was carried out by myself. The TmPPase:Mg_5IDP crystals had been grown via the sitting-drop vapour diffusion technique at room temperature in 32% PEG 400, 0.1 M Tris (pH 8.5), 0.1 M $MgCl_2$, 0.1 M NaCl and 2 mM DTT. The TmPPase:$MgWO_4$ crystals had been grown at 4 ˚C using the sitting-drop vapour diffusion technique in 19-22 % PEG 350 monomethylether, 0.1 M MES-NaOH pH 6.5, 0.2 M $CaCl_2$ and 2 mM DTT before being soaked in 10 mM Na_2WO_4. Datasets were collected on beamline i04 at Diamond Light Sourse (DLS) for TmPPase:Mg_5IDP and id23-1 at ESRF for TmPPase:$MgWO_4$. Initial data processing was carried out in XDS to obtain merged and scaled data. Molecular replacement was carried out in Phenix.Phaser using the high resolution VrPPase:Mg_5IDP structure as a search model (PDB ID: 4A01) with iterative rounds of refinement carried out in Phenix.refine using coot to make manual alterations to the model (for full refinement information, see methods).

Following refinement, the R-factors for TmPPase:Mg_5IDP were 24.8 %/27.5 % with 95.4 % of residues in the preferred region of the Ramachandran plot. The TmPPase:$MgWO_4$ structure had R-factors of 23.2 %/27.9 % with 93.4 % of residues in the preferred regions of the Ramachandran plot. Full data collection and refinement statistics can be found in (Table 4.1).

Table 4.1: Data Collection and Refinement Statistics for TmPPase Structures.

Crystal	**TmPPase:Mg_5IDP**	**TmPPase:$MgWO_4$**
Data collection		
Space group	P $2_12_12_1$	P 2_1
Cell dimensions		
a, b, c (Å)	106.4, 106.8, 161.9	83.7, 108.9, 105.8
β (˚)	90	108.8
Source	DLS-i04	ESRF ID23-1
Wavelength (Å)	0.9795	1.214
Resolution (Å)*	30-3.49 (3.70-3.49)	25-4.0 (4.1-4.0)
R_{merge} (%)*	9.3 (140.8)	5.5 (42.6)
I/δ*	15.83 (1.35)	14.3
Completeness (%)*	99.5 (98.6)	98.8 (92.8)
Redundancy*	8.8 (8.6)	3.8 (3.4)
Refinement		
Resolution (Å)	3.5	4.0
No. of reflections	23,904	29,597
R_{work}(%)/R_{free}(%)	24.3/27.3	23.2/27.9
No. of atoms	10,443	9,808
Protein	10,389 (1,450 residues)	9,794 (1,383 residues)
IDP or WO_4	36 (2 IDP)	10 (2 WO_4)
Mg^{2+}	10	2
Na^{+}	2	-
No. of chains/ASU	2	2
B-factors (Å^2)		
All atoms	131.0	134.9
Protein	130.9	133.0
IDP or WO_4^{3+}	167.8 (IDP)	198.5 (WO_4^{3+})
Mg^{2+}	134.9	151.6
Na^{+}	112.7	-
r.m.s.d.		
Bond length (Å)	0.005	0.004
Bond angle (˚)	0.70	1.19

*Values in parenthesis are for the highest resolution shell.

4.2.2. Analysis of TmPPase:Mg_5IDP and TmPPase:$MgWO_4$ Structures

There was clear, continuous and unambiguous density for the helices in each model (Figure 4.1). There was additional density for many loop regions in the TmPPase:Mg_5IDP structure, including the 5-6 loop over the active site. These loops were omitted from the TmPPase:$MgWO_4$ structure, along with the side chains of many longer residues, such as lysine and arginine, due to a high degree of disorder in these regions resulting in no electron density for these entities.

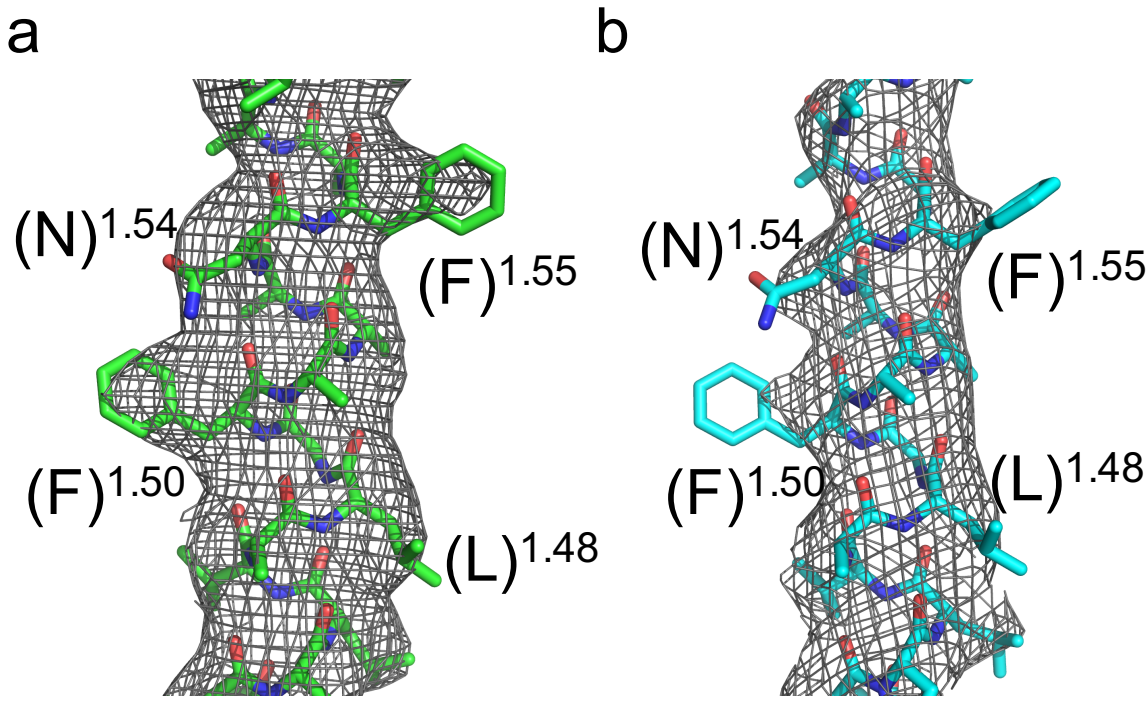


Figure 4.1: Representative Electron Density of TmPPase Structures.
F_o-F_c density for a section of helix 1 representative of helical density within (a) TmPPase:Mg_5IDP at 3.5 Å resolution and (b) TmPPase:$MgWO_4$ at 4.0 Å resolution. Residues are labelled in parenthesis as they are not conserved between M-PPases of different species.

In addition, there was positive density from an F_o-F_c map (3 δ) in the active sites for Mg_5IDP and $MgWO_4$ in each model, respectively (Figure 4.2). The position of the Mg_5IDP complex very closely overlays with the same complex found in the previous VrPPase:Mg_5IDP structure (PDB ID: 4A01) (Figure 4.3a), with a similar closure of the active site helices when compared with the resting state TmPPase structure (PDB ID: 4AV3) (Figure 4.3b). However, given the modest resolution of these structures (3.5 – 4.0 Å), the position of the ligands is based on the unbiased positive F_o-F_c density but with comparisons made between the new structures and the position of the ligands in the

previous solved, higher resolution structures. This was seen by the movement of helices 5, 6 and 11 between the TmPPase:CaMg and TmPPase:Mg_5IDP structures (Figure 4.3b) to constrict the active site, a movement not seen between the TmPPase:CaMg structure and the TmPPase:$MgWO_4$ structures (Figure 4.4). In the absence of loop 5-6 in the TmPPase:$MgWO_4$ structure, the channel was open on the cytoplasmic side of the active site and there was very little interaction seen between the $MgWO_4$ complex and the surrounding residues (Figure 4.2b, Figure 4.4). This was due to the absence of visible water molecules, which would likely be filling the remainder of the active site but which cannot be fit to the density at this resolution (4 Å).

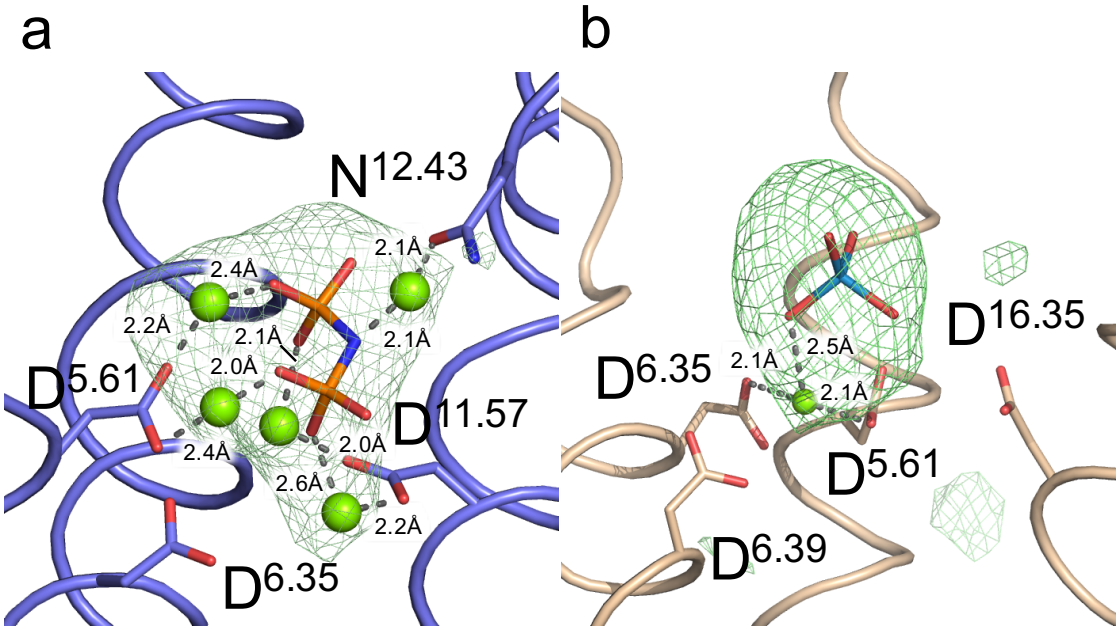


Figure 4.2: F_o-F_c Density at the Hydrolytic Centre of TmPPase Structures.
Positive F_o-F_c electron density (3 δ) for (a) Mg_5IDP complex in TmPPase:Mg_5IDP and (b) $MgWO_4$ in TmPPase:$MgWO_4$. Mg_5IDP and $MgWO_4$ are superimposed over unbiased positive F_o-F_c observed prior to their addition to the model. Figure panels reused with permission (Supplementary Figure 2, (K. M. Li et al., 2016)).

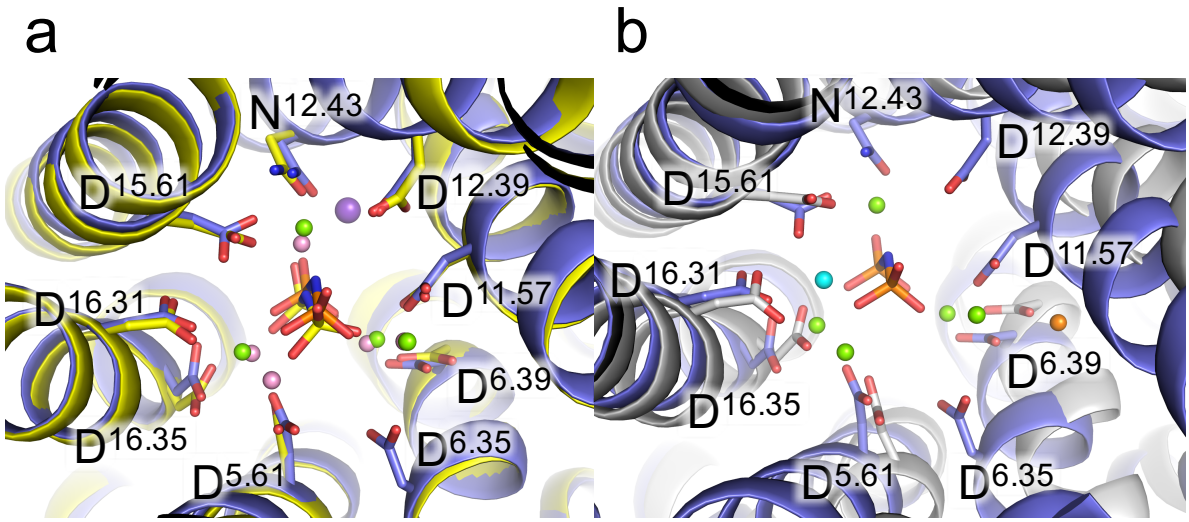


Figure 4.3: Hydrolytic Centre Overview of TmPPase:Mg_5IDP.
Comparison of the active site residues between (a) TmPPase:Mg_5IDP (blue carbon, orange IDP, green Mg^{2+}) and VrPPase:Mg_5IDP (yellow carbon, yellow IDP, salmon Mg^{2+}) and (b) TmPPase:Mg_5IDP and TmPPase:CaMg (grey carbon, cyan Ca^{2+}, orange Mg^{2+}).

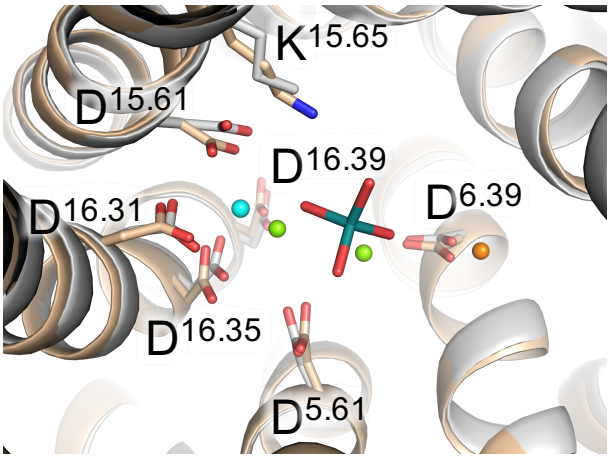


Figure 4.4: Hydrolytic Centre Overview of TmPPase:$MgWO_4$
Comparison of the active site residues between TmPPase:$MgWO_4$ (beige carbon, marine WO_4^{3+}, green Mg^{2+}) and TmPPase:Ca:Mg (grey carbon, cyan Ca^{2+}, orange Mg^{2+}).

There was strong positive density (F_o-F_c, 4.7 δ) around the ion gate in protomer A of TmPPase:Mg_5IDP (Figure 4.5). A sodium ion was refined in this position and the resulting density fits this placement. The density alone was insufficient evidence to suggest a sodium ion, since water molecules also have similar electron density. The resolution was also insufficient to accurately identify this density as a sodium ion. Alternative methods were used to identify and correctly assign this density as a sodium ion. The ion was penta-coordinated, with surrounding residues at distances between 2.39-2.49 Å. The results of a valence test (Section 2.2.5.4) carried out at this position also has a score of 0.9, with scores over 0.7 showing a high probability of being a sodium ion. In

contrast, tests carried out on the magnesium ions in the active site and water molecules elsewhere in the structure (in positions known to be correct based on previous structural data) showed relevant scores for magnesium ions and water molecules, validating the presence of sodium ions at this position. Moreover, previous mutational studies on several residues in proximity to the ion gate found that $D^{16.46}$ was essential for Na^+-binding and activity of the protein (Belogurov *et al.*, 2005).

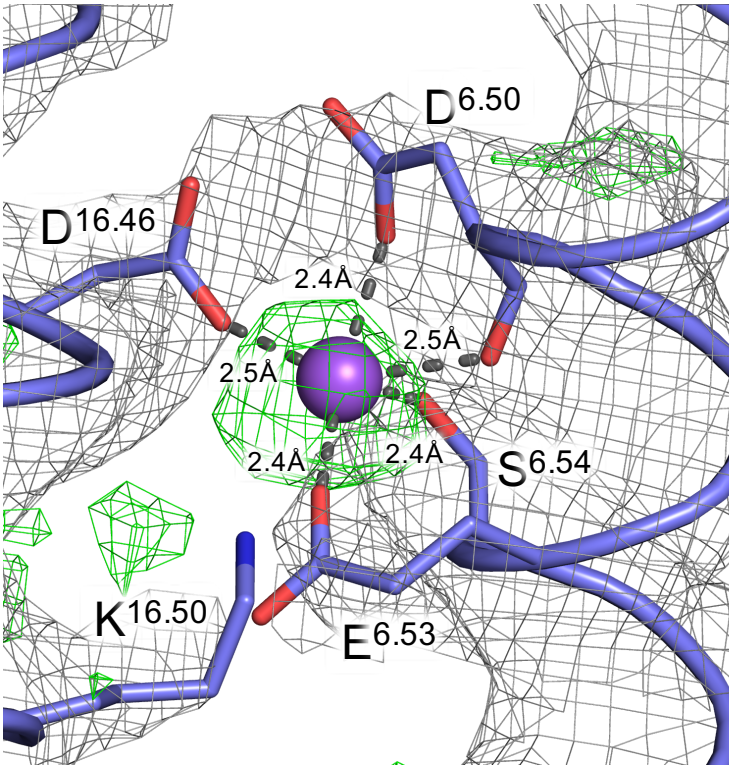


Figure 4.5: Coordination of Na^+ at the Ion Gate of TmPPase.
Positive F_o-F_c electron density for Na^+ near the ion gate in TmPPase:Mg_5IDP at 3 δ. $2F_o$-F_c density is also shown for surrounding helices (1 δ). Na^+ (purple sphere) is superimposed onto the unbiased positive F_o-F_c density and is shown prior to its addition to the model. Figure adapted with permission (Supplementary Figure 2, (K. M. Li et al., 2016)).

4.2.3. Identifying the Sodium Ion Binding Site

The proposed sodium ion binding site in TmPPase:Mg_5IDP consists of residues $D^{6.50}$, $E^{6.53}$, $S^{6.54}$ and $D/N^{16.46}$ (Figure 4.6a). There were three noticeable differences between this structure and VrPPase:Mg_5IDP (Figure 4.6b). The first was the position of $E^{6.53}$ in TmPPase, which is located one helical turn down in VrPPase ($E^{6.57}$). The second is one of the ion gate residues, $D/N^{16.46}$, which is an aspartate in TmPPase but an asparagine in almost all other M-PPases, including VrPPase. The third is the position of the lysine ($K^{16.50}$). In VrPPase:Mg_5IDP, the NH_3 group of $K^{16.50}$ occupies the same position as the

sodium ion in TmPPase:Mg_5IDP. It was previously thought the switch of glutamate position accounted for differences in the binding and pumping of sodium ions and protons (Tsai *et al.*, 2014). However, this glutamate is located at position $E^{6.53}$ for both Na^+-PPases and K^+-independent H^+-PPases, which implies the specific mechanisms of ion selectivity are still not known and are indeed complex.

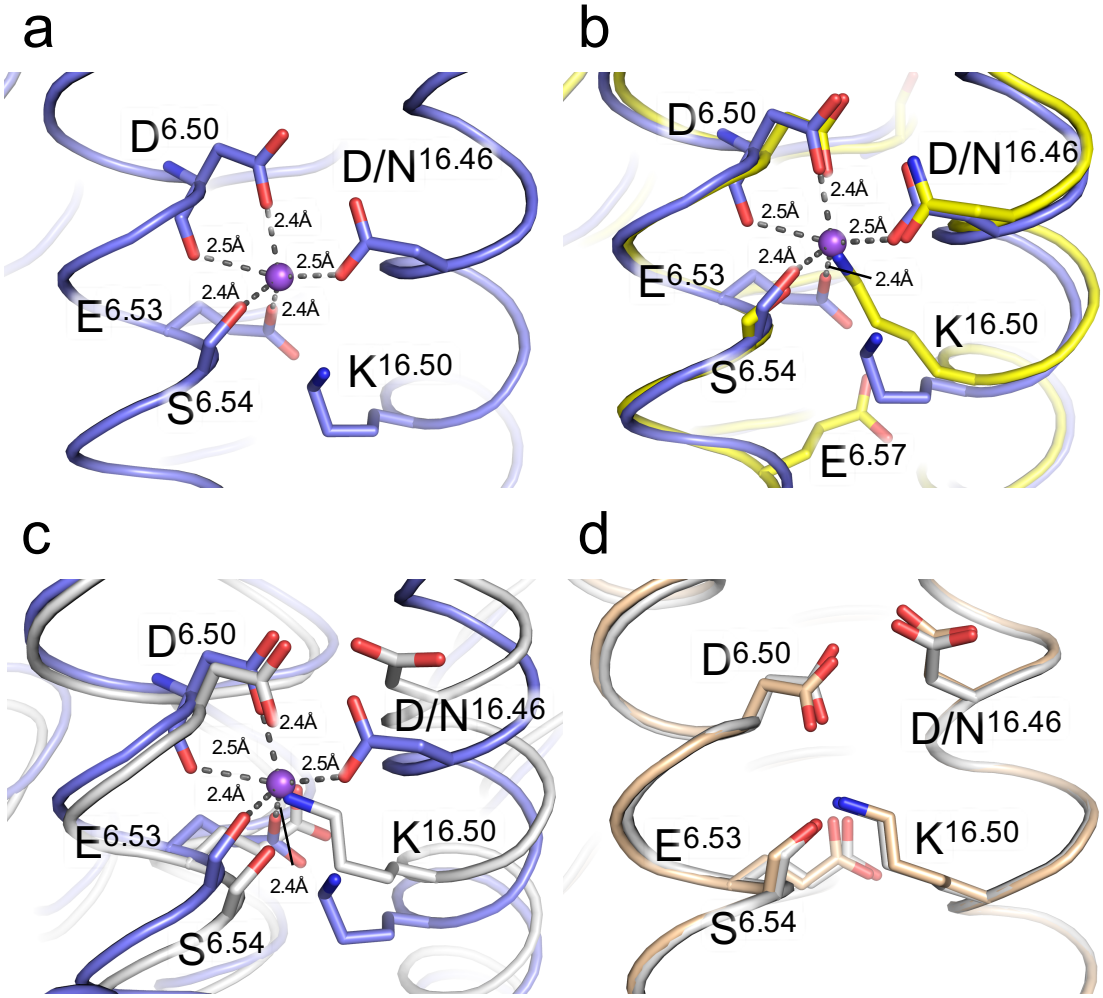


Figure 4.6: Ion Gate Overview of TmPPase Structures.
Key ion gate residues for (a) TmPPase:Mg_5IDP (blue carbon, purple Na^+) and comparisons with (b) VrPPase:Mg_5IDP (yellow), (c) TmPPase:CaMg (grey). (d) Comparison of ion gate residues between TmPPase:$MgWO_4$ (beige) and TmPPase:CaMg (grey).

The displacement of $K^{16.50}$ was observed when comparing the TmPPase:Mg_5IDP structure with the resting state TmPPase structure (TmPPase:CaMg) (Figure 4.6c). As with the VrPPase:Mg_5IDP structure, the amine group of $K^{16.50}$ occupies the position of Na^+ in the TmPPase:Mg_5IDP structure. $D^{16.46}$ has also changed conformation in the resting state structure, moving in to coordinate the sodium ion upon binding. The position

of the ion gate residues in TmPPase:$MgWO_4$ can be superimposed with those from TmPPase:CaMg.

4.2.4. Comparison Between Previous and New M-PPase Structures

The TmPPase:Mg_5IDP structure very closely compared with the previous VrPPase:Mg_5IDP structure (r.m.s.d. = 0.9 Å, comparing Cα of all residues) and showed perfect superimposition of the active site residues and overall secondary structure. The main differences were in the position of the ion gate residues brought about by minor sequence differences and the presence of a sodium ion in the TmPPase:Mg_5IDP structure, which had altered the interactions of the ion gate residues between these structures. The TmPPase:Mg_5IDP structure deviated more from the TmPPase:CaMg structure (r.m.s.d. = 1.6 Å, comparing Cα of all residues), due to major differences upon substrate or inhibitor binding. The closing of the active site loops and constrictions of the helices around this region led to major conformational changes in the helices. Additional changes near the ion gate, such as the displacement of $K^{16.50}$, also plays an important role in forming the sodium binding site.

Inversely, the TmPPase:$MgWO_4$ structure more closely resembles the resting state TmPPase structure (r.m.s.d. = 0.4 Å, comparing Cα of all residues) than the TmPPase:Mg_5IDP structure (r.m.s.d. = 1.7 Å, comparing Cα of all residues). Both the active site and ion gate are highly similar between TmPPase:$MgWO_4$ and TmPPase:CaMg, with almost no difference between the residues and helices at these positions. This was likely due to WO_4^{3+} being an analogue of the product-bound form of TmPPase (P_i). During the catalytic cycle this would be the last product released prior to reaching the resting-state and therefore the sodium ion would already have been pumped. This would leave the ion gate residues in the same conformation as the resting state structure and would also ensure that the active site loops had moved away from the active site and become disordered to allow for product release.

4.3. Structural Studies of PaPPase Using X-ray Crystallography

4.3.1. Crystallisation Trials of PaPPase

Initial crystallisation trials of PaPPase were carried out in sparse matrix screens with protein in DDM and DM. Multiple 96-well plates were set up containing commercially

available screens, such as MemGold 1 and 2 as well as MemStart & MemSys using the sitting-drop vapour diffusion method. Very few crystals were obtained from these screens and these diffracted to around 15-20 Å, which was insufficient to obtain a correct indexing solution or to gain any useful structural information. Alongside these trials, a new optimised crystal screen was set up based upon hits obtained in previous studies (Kellosalo *et al.*, 2013). The optimised screen covered conditions from the best crystal hit (35 % PEG 400, 0.1 M Tris pH 8, 0.1 M KCl), which had yielded crystals diffracting to 5.8 Å. Full details of this screen can be found in the materials and methods section, but essentially involved varying the concentration and pH of different buffers and PEG 400 alongside trying alternative salts and salt concentrations.

4.3.2. Initial Crystallisation Hits of PaPPase and 6.5 Å Dataset

4.3.2.1. Collection of Initial PaPPase Datasets

Using the optimised screen yielded numerous crystals of needle and plate morphologies that diffracted to < 10 Å, with the best hit from the initial crystal screens found at 6.5 Å with the protein in DM. This crystal was grown in 30 % PEG 400, 0.1 M MES pH 6.5, 0.05 M $LiSO_4$ and 0.05 M NaCl using the sitting drop vapour-diffusion method (Figure 4.7). A full dataset was collected and indexed into the P 2_1 spacegroup (Table 4.2).

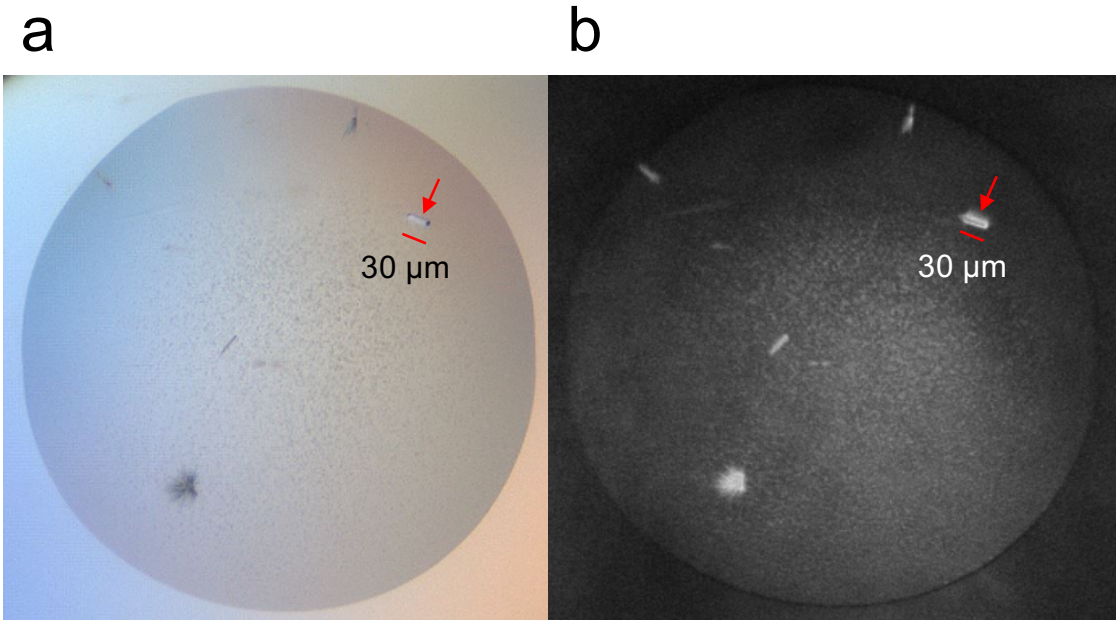


Figure 4.7: Morphologies of 6.5 Å diffracting PaPPase crystals.
Size and morphology of PaPPase crystals diffracting to 6.5 Å using (a) standard imaging and (b) UV-TPEF imaging, where crystals showing UV fluorescence appear white. Red arrows indicate the crystal used to obtain the data set and resulting structure.

Table 4.2: Data Collection Statistics for 6.5 Å PaPPase Structure.

Crystal	**PaPPase (6.5 Å)**
Data collection	
Space group	P 2_1
Cell dimensions	
a, b, c (Å)	106.6, 87.0, 116.3
β (˚)	108.4
Source	
Wavelength (Å)	0.966
Resolution (Å)	48.7 - 6.5 (7.4 - 6.5)
CC(1/2) (%)*	94.7 (21.8)
I/δ*	1.79 (0.38)
Completeness (%)*	95.5 (80.4)
Redundancy	3.2 (3.1)

*Value in parenthesis are for the highest resolution shell.

4.3.2.2. Processing of Initial PaPPase Dataset

At this resolution, it was possible to solve the structure by molecular replacement using a homology model generated by Sculptor (part of the Phenix software package) based on the TmPPase resting state structure using the sequence of PaPPase. This model lacked the catalytically important loop regions (loop 5-6), all water molecules and metal ions and all residues had been truncated to alanine as side chain density would not be visible at this resolution. The resulting structure showed a dimer, with each protomer containing 16 transmembrane helices (Figure 4.8). The helical positions most closely matched those of the IDP-bound forms of M-PPases, but no positive density (F_o-F_c) could be seen in the active site, nor was it possible to see density for the 5-6 loop or any side chains throughout the protein. Due to the limited resolution, the structure was not solved and no further refinement was carried out following the initial observation of the electron density maps.

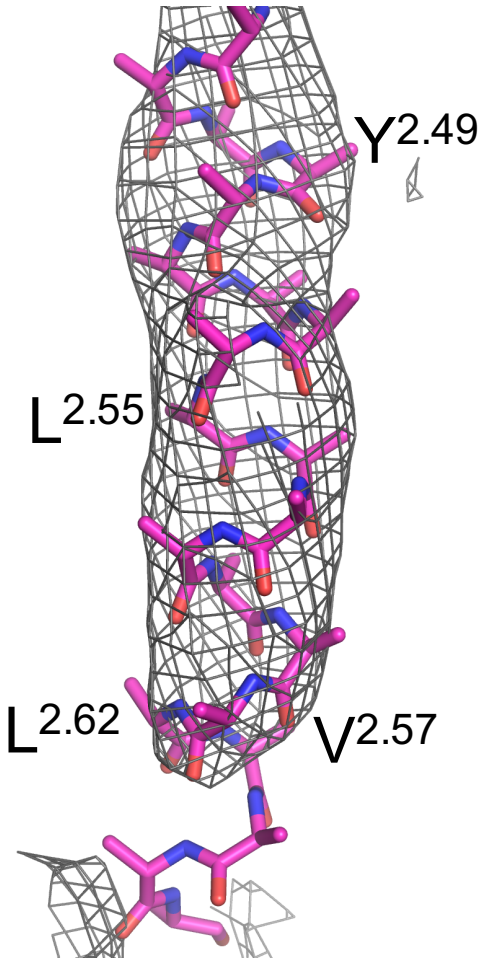


Figure 4.8: Representative Electron Density of 6.5 Å PaPPase Structure.
Representative $2F_o$-F_c density for helix 2 of the 6.5 Å PaPPase structure. Side chains have not been added to the model due to the lack of density at this resolution.

4.3.2.3. Structural Observations of the 6.5 Å PaPPase Structure

Despite the low-resolution electron density maps, it was possible to observe crystal packing for this structure (Figure 4.9). It appears that the contacts form more closely in two directions, with the cytoplasmic side of protomer A forming contacts with the cytoplasmic side of protomer B from the neighbouring dimer. This explains the crystal morphology, since the lack of contacts on one axis leads to poor crystal formation along this axis, resulting in the formation of plate and needle crystal morphologies. This third axis is likely weaker at forming contacts due to the position of the detergent micelles, which are found around this portion of the protein. Therefore, it may be possible to switch to a detergent with a smaller micelle size or to try reducing the concentration of detergent to improve crystal formation. Moreover, dehydration experiments may lead to tighter crystal contacts and thus better crystal formation and diffraction. Contacts between the protomers at the exterior side of the channel were also not visible, as the poor crystal

contacts had left the side chains disordered, generating poor electron density at these regions.

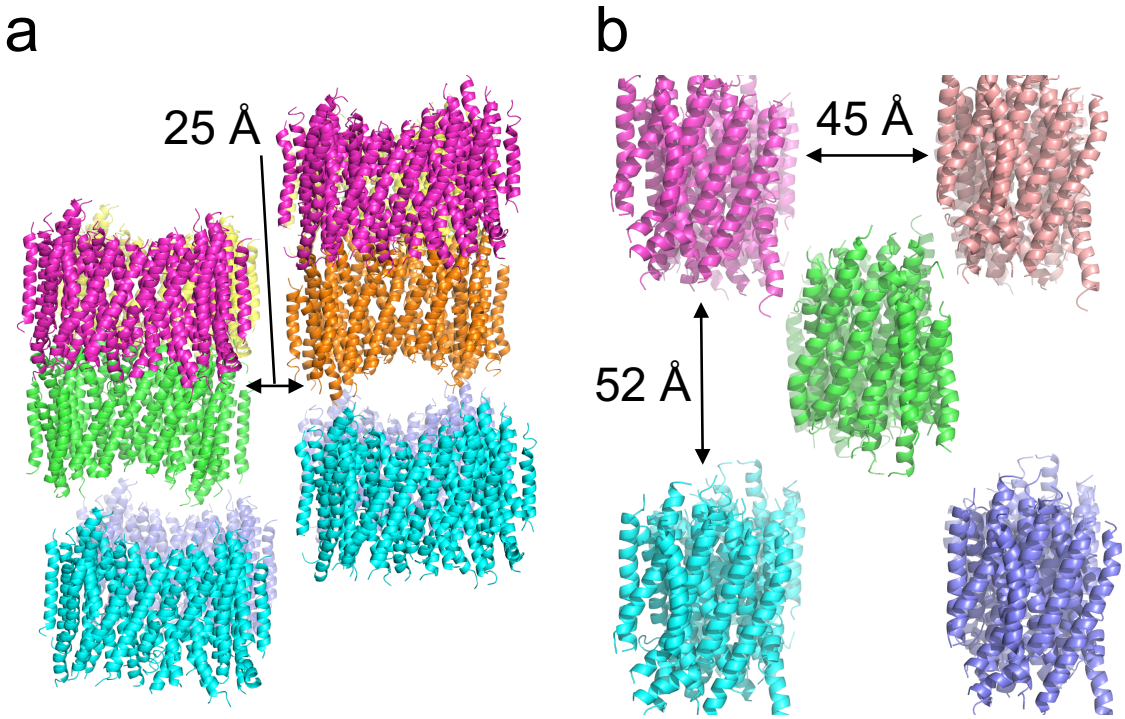


Figure 4.9: Crystal Packing of the 6.5 Å PaPPase Structure.
Crystal packing of the 6.5 Å PaPPase dataset shown (a) along the x-axis and (b) along the y-axis.

4.3.3. Optimisation of PaPPase Crystals to Improve Diffraction

4.3.3.1. Optimisation of PaPPase Crystallisation Conditions

Following on from the initial 6.5 Å dataset, hundreds of crystals were obtained and tested, leading to dozens of additional datasets at the same resolution. Different detergents were tested, including DDM, DM, NM, OM, UM and C8E4 at 10 - 20 x CMC, but none of these produced crystals diffracting better than the initial 6.5 Å hit. Multiple other detergents had been tested in previous studies to see what effect these would have on crystallisation, including OG, CHAPSO, Cymal-5 and 6 and also OGNG, but none of these led to better diffracting crystals.

The type of detergent used was likely not the main deciding factor on how the crystals formed or how well they diffracted. Likewise, the well conditions of the optimised screen were varied, including different concentrations of buffer, pH as well as the type and concentration of salts. Moreover, the original sparse matrix screens had yielded no additional hits than those found in the original crystallisation studies of PaPPase. Thus,

it seemed that alternative methods of optimisation were required to further improve the resolution.

4.3.3.2. "HiLiDe" Optimisation Technique

One such optimisation method was the "HiLiDe" method (High Lipid and Detergent), which involved the addition of increasing concentrations of lipids and detergents in a grid to find an optimal combination that could significantly improve the growth and diffraction of crystals (Section 2.2.2.1). Initial screens were based on conditions from the best diffracting crystals (30 % PEG 400, 100 mM MES, pH 6.5, 50 mM $LiSO_4$ and 50 mM NaCl). Initial trials involved increasing concentrations of DDM between 1 and 6 % alongside increasing PEG 400 concentrations from 30-38 %. Other plates involved increasing concentrations of DM and NM between 0 and 3 %. Crystals obtained from this method were of a thin needle morphology (~10 μm width) and did not diffract better than 15 Å. Indexing was unsuccessful given the lack of clear spots on the diffraction images and so no additional information could be gained from these crystals. Given the lack of any improved crystals, dehydration experiments were set up to try and improve crystal formation and diffraction.

4.3.3.3. Dehydration Experiments of PaPPase Crystals

The additional 96-well crystallisation plates set up using the PaPPase optimisation screen reliably produced crystals that diffracted to 6.5 Å, which were used for a series of dehydration experiments. Dehydration experiments can sometimes improve crystal diffraction by reducing the water content of the crystals, creating better crystal packing and thus stabilising the crystal, leading to better diffraction. I used two main methods. The first was the transference of fully formed crystals into solution containing the mother liquor with added concentrations of glycerol and/or PEG 400. The second was to add glycerol or PEG 400 directly to the well solution. In both cases the hope was to dehydrate the crystal or the drop and allow it to form a more tightly packing crystal that would show improved diffraction. However, neither method improved the diffraction of the crystals as the best one obtained still diffracted to around 6.5 Å.

4.3.3.4. Reduced Detergent Concentration in Crystallisation of PaPPase

Another method of optimisation was to try and reduce the concentration of detergent present at the point of crystallisation. This method would lower the free detergent concentration as well as lower the number of empty detergent micelles in the sample, reducing the number of detergent crystals whilst increasing the number of protein-protein crystal contacts, improving diffraction of the resulting crystals. This was achieved by adding lower concentrations of detergent during the detergent exchange step of purification from 20 x CMC to 4 x CMC. In doing so, this would reduce the final concentration of detergent in the protein sample following the protein concentration step of purification. However, as with other optimisation methods attempted, multiple crystals were collected and tested but there was no improvement in crystal diffraction as data sets obtained were at 6.5 Å or worse.

4.3.4. Improved PaPPase Datasets and 4.3 Å Structure

4.3.4.1. Crystallisation Conditions and Data Collection of 4.3 Å Dataset

During data collection, one crystal was found that diffracted to 4.3 Å. I noted that most of the solvent in the loop was absent, effectively dehydrating the crystal. This crystal had been grown in 33 % PEG 400, 0.1 M MES pH 6.5, 0.05 M $LiSO_4$ and 0.05 M NaCl via the sitting drop vapour diffusion method (Figure 4.10). The crystal in question had been grown with several dozen more as part of the dehydration experiments, but was a control crystal for comparison to the dehydrated crystals. The crystallisation conditions themselves were only marginally different from the ones used to obtain the 6.5 Å dataset (33 % PEG 400 instead of 30 % PEG 400) and indeed these conditions had been tested previously, with only 6.5 Å-diffracting crystals found. The only difference to previous attempts was that the solvent in the loop had mostly evaporated, leaving the crystal in dehydrating conditions.

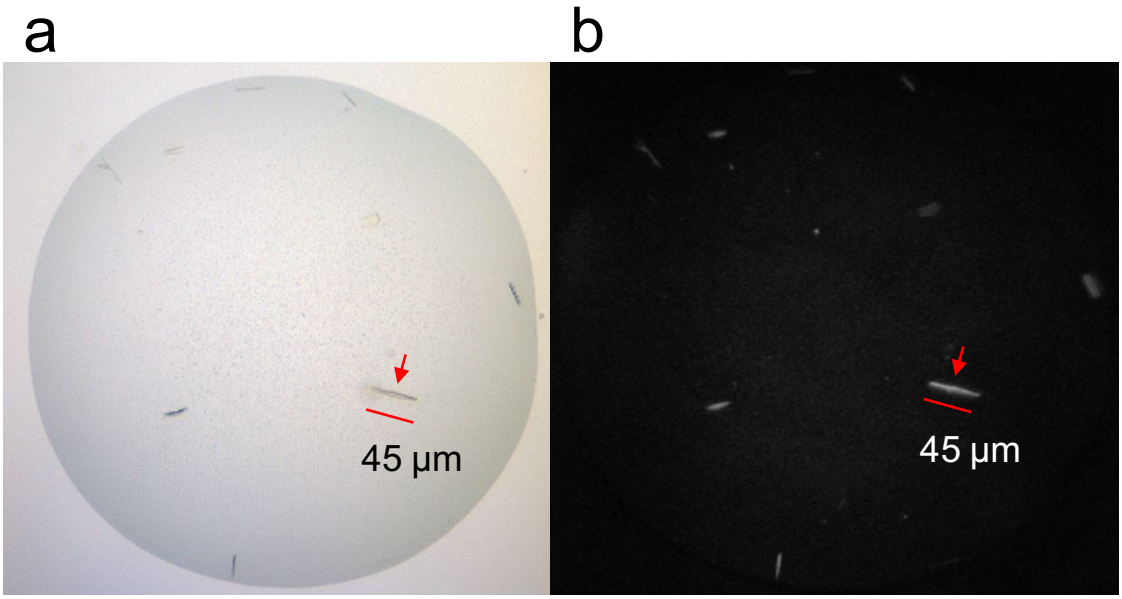


Figure 4.10: Morphology of 4.3 Å diffracting PaPPase Crystals.
Size and morphology of PaPPase crystals diffracting to 4.3 Å using (a) standard imaging and (b) UV-TPEF imaging, where crystals showing UV fluorescence appear white. Red arrows indicate the crystal used to obtain the data set and resulting structure.

4.3.4.2. Data Processing and Structural Refinement of 4.3 Å Dataset

A full data set was collected from this crystal (Table 4.3). The data were processed in XDS and a molecular replacement solution was found using a PaPPase homology model based on the TmPPase:Mg_5IDP structure with flexible loop regions removed. The data underwent limited initial refinement, with the R-factors being 37.8%/43.8%, but a full refinement was not carried out due to the collection of higher resolution data (Section 4.3.5). The overall helical density was much improved compared with the 6.5 Å dataset (Figure 4.11). At this resolution, electron density was visible for some of the side chains and allowed for more accurate fitting of the helices compared with the 6.5 Å structure. Despite this, there was still missing density, particularly around the active site. No density could be seen for the Mg_5IDP complex or for loop 5-6, which was observed packing over the active site in both the TmPPase and VrPPase IDP-bound structures.

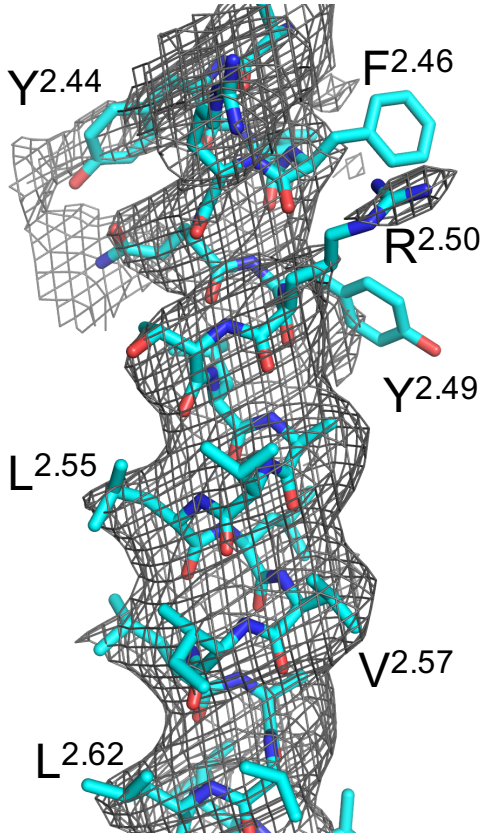


Figure 4.11: Representative Electron Density for 4.3 Å PaPPase Structure.
Representative ($2F_o$-F_c) density (1 δ) for helix 2 of the 4.3 Å PaPPase structure.

In the 4.3 Å PaPPase structure, it was noted that the final 11 residues of the C-terminus were absent from the model due to a lack of electron density in these regions (Figure 4.12a). This was in contrast to the TmPPase structures (Figure 4.12b), in which these residues are all visible and form vital interactions across the dimer interface, maintaining the overall structure of TmPPase. The lack of electron density for these residues implied a degree of disorder in this region of the protein, but the C-terminal truncation mutant of PaPPase (Section 3.6) showed a significant decrease in hydrolytic activity compared with the wild-type protein. Therefore, despite the lack of electron density, these regions were clearly still important to the overall structure and function of these proteins, particularly during solubilisation at 70 ˚C. Another explanation for the lack of electron density is the possibility that the expressed PaPPase protein lacks the C-terminal residues, since the purification gel shows another band directly above the main purification band (Figure 3.12). However, this is unlikely to be the case, since sequence-confirmed expression and purification of the C-terminal truncation mutant of PaPPase was almost entirely inactive (10-fold lower than the wild-type PaPPase. If the C-terminal residues were missing, the activity would decrease and no difference would be seen between the wild-type protein

and the C-terminal truncation mutant. It is therefore more likely that the residues are either disordered or not resolvable at this resolution.

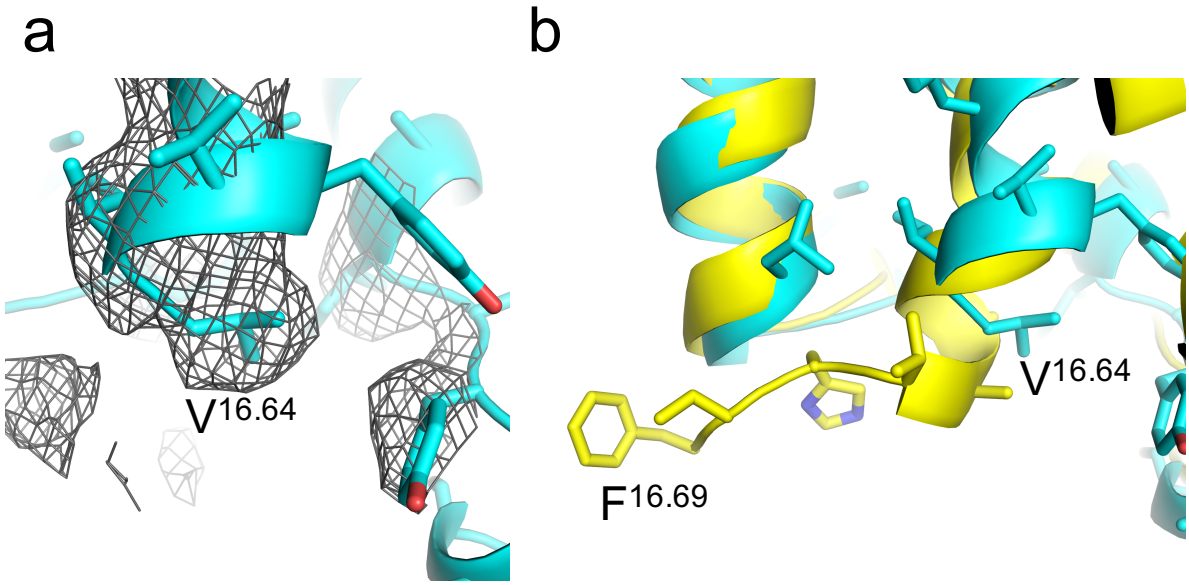


Figure 4.12: Overview of the PaPPase C-terminus.
(a) $2F_o$-F_c electron density (3 δ) for the C-terminal of the 4.3 Å PaPPase structure showing the absence of density for the final 13 residues and (b) comparison between the C-terminal residues of the 4.3 Å PaPPase structure (cyan) with the TmPPase:CaMg structure (yellow).

Table 4.3: Data Collection and Refinement Statistics of the 4.3 Å PaPPase Structure.

Crystal	**PaPPase (4.3 Å)**
Data collection	
Space group	P 2_1
Cell dimensions	
a, b, c (Å)	107.5, 88.5, 117.1
β (˚)	107.2
Source	
Wavelength (Å)	0.966
Resolution (Å)	47.5 - 4.3 (4.5 - 4.3)
CC(1/2) (%)*	98.9 (29.9)
I/δ*	3.9 (0.6)
Completeness (%)*	99.1 (98.9)
Redundancy	3.4 (3.4)
Refinement	
Resolution (Å)	4.3
No. of reflections	48,261
R_{work}(%)/R_{free}(%)	37.8 / 43.8
No. of atoms	9,440
Protein	9,440
IDP	0
Mg^{2+}	0
No. of chains/ASU	2
B-factors (Å^2)	
All atoms	29.97
Protein	29.97
Ligand	X
r.m.s.d.	
Bond length (Å)	0.01
Bond angle (˚)	1.68

*Values in parenthesis are for the highest resolution shell.

4.3.5. Merging of Multiple PaPPase Datasets and the 3.8 Å structure

4.3.5.1. Data Collection from Multiple PaPPase Crystals

Once crystals had been found that diffracted to 4.3 Å, I could repeatedly obtain these crystals by dehydrating them during crystal picking. This was achieved by leaving the crystals in the air for 10-30 seconds once picked prior to flash-freezing in liquid nitrogen. I screened dozens of crystals using the 96-well optimisation plates and from this I obtained 7 full datasets each between 4.0 and 4.5 Å resolution. In each case the space group and unit cell parameters were almost identical (Table 4.4). Molecular replacement and a single round of refinement was carried out to obtain electron density maps. All structures showed identical crystal packing and positive density (F_o-F_c) in the active site at the position of the Mg_5IDP complex in the TmPPase and VrPPase IDP-bound structures.

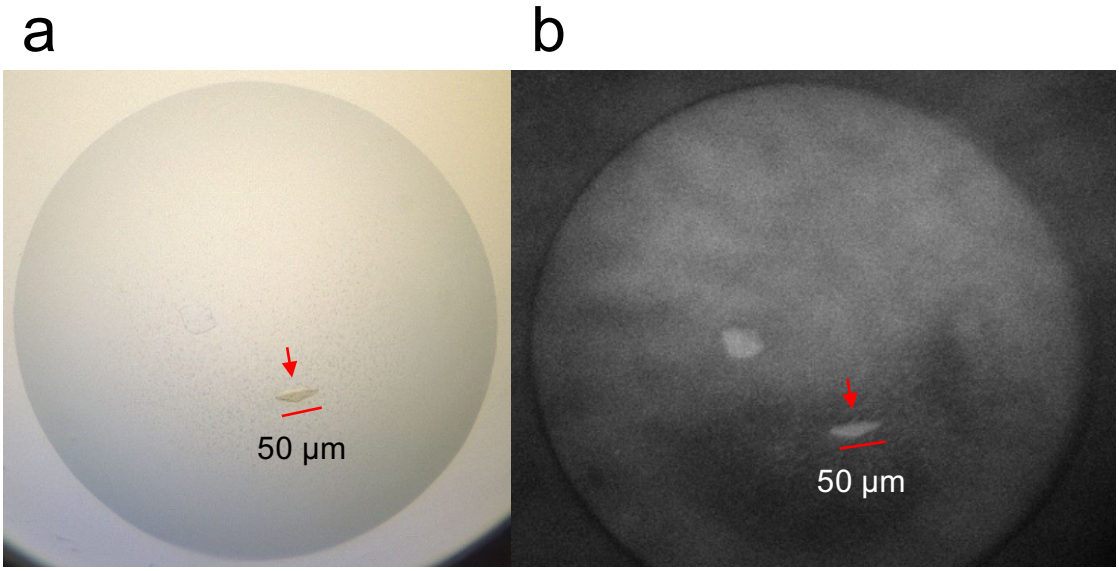


Figure 4.13: Morphology of Crystals used to Generate the 3.8 Å PaPPase Structure.
Size and morphology of one of the PaPPase crystals containing data to 3.8 Å using (a) standard imaging and (b) UV-TPEF imaging, where crystals showing UV fluorescence appear white. Red arrows indicate the crystal used to obtain the data set that was used during merging of data and used to solve the structure.

4.3.5.2. Data Processing and Merging of PaPPase Datasets

During data processing in XDS, it was noted that the first few hundred images (10-40˚) of each dataset contained higher resolution data (~ 3.8 Å) but with a lower overall completeness (~40-70 %). The datasets were reprocessed using only a select number of images (Table 4.4). Since each dataset had identical space groups and very similar unit cell parameters, it was possible to combine these partial datasets. Datasets were merged using XScale and the merged dataset could be cut at a resolution of 3.8 Å without significant loss of completeness and with a CC(1/2) of around 30 % (Table 4.5).

Table 4.4: Data Collection and Processing Statistics from Multiple PaPPase Crystals.
Data processing statistics for truncated datasets of PaPPase used to create combined dataset used to obtain 3.8 Å structure.

Dataset	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Beamline	DLS-i04	DLS-i04	DLS-i04	DLS-i04	ESRF MASSIF-1	ESRF ID23-1	DLS-i24
Image Range (0.1˚ / image)	1-400	1-180	1-200	1-150	1-180	1-300	1-200
Spacegroup	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
a (Å)	109.5	113.3	112.1	109.6	110.1	111.1	106.9
b (Å)	89.2	89.9	89.2	88.7	90.9	89.5	88.7
c (Å)	118.9	119.4	117.8	116.8	121.1	120.4	116.6
α (˚)	90	90.1	90.1	90.1	91.1	90.3	89.9
β (˚)	106.7	107.2	107.3	107.5	107.2	106.7	107.1
γ (˚)	90	90.1	90	90.4	89.7	89.9	90.6
Resolution (Å)	3.96	4.35	4.12	3.99	4.01	3.94	3.84
Completeness (%)*	67.6 (68.6)	49.5 (46.5)	51.6 (50.3)	39.3 (39.1)	52.1 (48.2)	45.9 (38.5)	48 (47.6)
I/δ*	3.6 (0.4)	4.2 (0.6)	4.9 (0.51)	2.5 (0.24)	2.7 (0.31)	3.8 (0.5)	2.5 (0.5)
CC(1/2) (%)*	99.8 (29.0)	99.9 (14.7)	100 (31.5)	99.9 (28.4)	99.7 (25.9)	99.5 (27.3)	99.7 (25.9)

*Values in parenthesis are for the highest resolution shell.

Table 4.5: Refinement Statistics of the Merged 3.8 Å PaPPase Structure.

Crystal	**PaPPase (3.8 Å)**
Data collection	
Space group	P 2_1
Cell dimensions	
a, b, c (Å)	107.2, 88.0, 116.8
β (˚)	107.0
Source	
Wavelength (Å)	0.96860
Resolution (Å)	20 - 3.8 (3.9 – 3.8)
CC(1/2) (%)*	97.5 (28.6)
I/δ*	3.6 (0.7)
Completeness (%)*	95.1 (82.5)
Redundancy	6.1 (3.9)
Refinement	
Resolution (Å)	3.8
No. of reflections	19,166
R_{work}(%)/R_{free}(%)	34.8 / 39.4
No. of atoms	9,486
Protein	9,456
IDP	18 (2 IDP)
Mg^{2+}	10
No. of chains/ASU	2
B-factors (Å^2)	
All atoms	117.2
Protein	117.2
Ligand	118.5
r.m.s.d.	
Bond length (Å)	0.003
Bond angle (˚)	0.685

*Values in parenthesis are for the highest resolution shell.

4.3.5.3. Refinement of the 3.8 Å PaPPase Structure

The merged 3.8 Å dataset was used for molecular replacement and refinement, and the structure was partially solved with R-factors of 34.8 %/39.4 % (Table 4.5). This showed a vast improvement in electron density over the 4.3 Å structure, with clearer density seen for the longer side chains, such as tyrosine, arginine and phenylalanine compared with the 4.3 Å structure (Figure 4.14). Despite the low resolution, more structural information could be observed compared with the 4.3 Å structure.

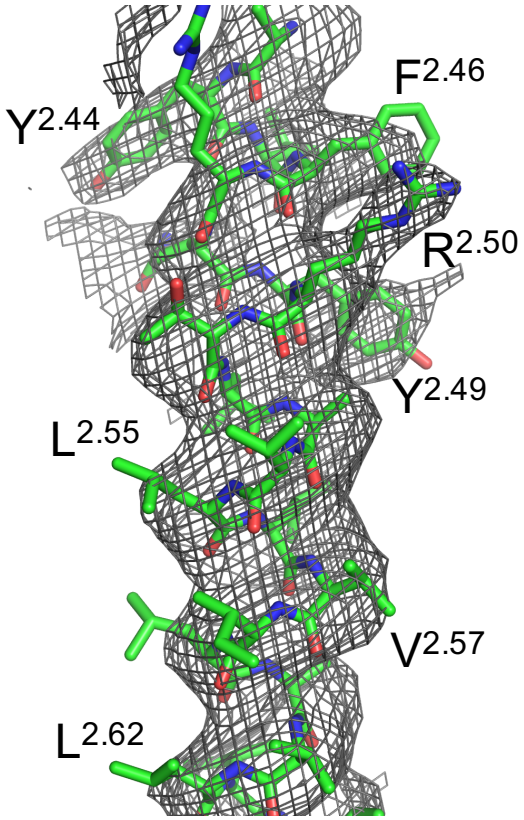


Figure 4.14: Representative Electron Density of the 3.8 Å PaPPase Structure.
Representative $2F_o$-F_c (3 δ) electron density for helix 2 of the 3.8 Å PaPPase structure.

4.3.5.4. Comparisons with Previous Structures at the Hydrolytic Centre

Electron density ($2F_o$-F_c) was observed in the active site at the position of the Mg_5IDP complex in both the TmPPase and VrPPase IDP structures (Figure 4.15a). This was added to the model and refined in place: the resulting electron density was able to accommodate the IDP and three of the magnesium ions, the remaining two magnesium ions did not produce negative density (F_o-F_c), but did not fit the $2F_o$-F_c density (Figure 4.15b). Comparison between the position of the Mg_5IDP complex in the 3.8 Å PaPPase

structure and VrPPase:Mg_5IDP showed that the position of the complex itself was quite similar, with only minor differences in the position of Mg^{2+} (Figure 4.16).

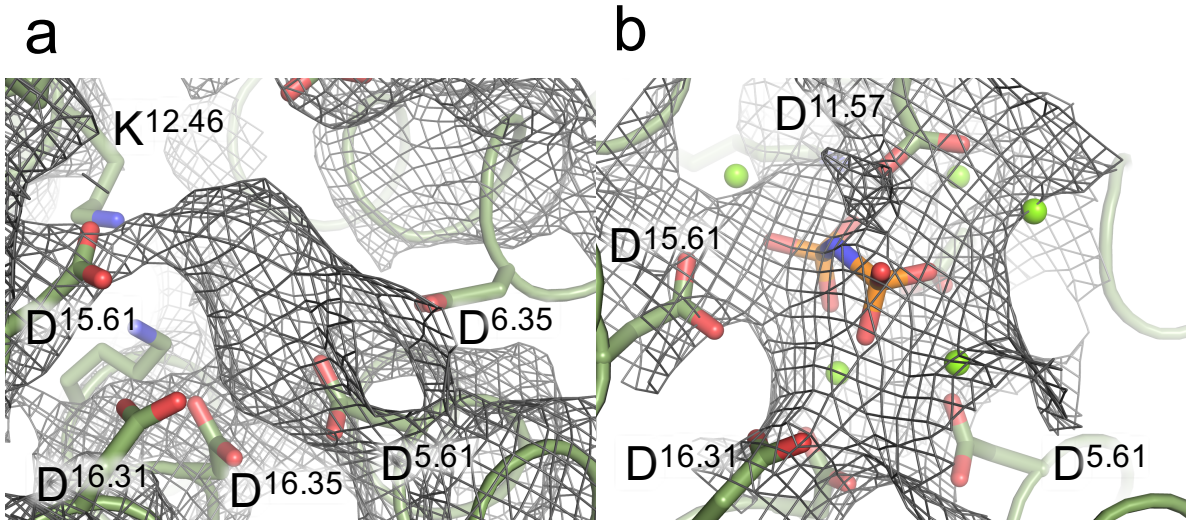


Figure 4.15: Electron Density at the Hydrolytic Centre of the 3.8 Å PaPPase Structure.
$2F_o$-F_c electron density (3 δ) for (a) the active site and surrounding residues of the 3.8 Å PaPPase structure following molecular replacement and (b) the active site of the same structure including Mg_5IDP complex at the end of refinement.

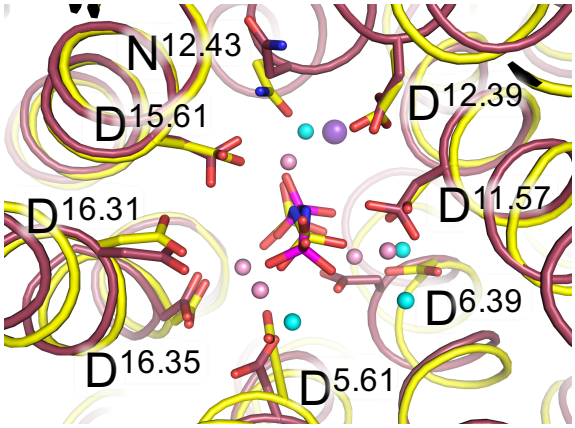


Figure 4.16: Hydrolytic Centre Overview of 3.8 Å PaPPase Structure.
Comparison between the hydrolytic centres of VrPPase:Mg_5IDP (yellow carbon, yellow IDP, salmon Mg^{2+}) and 3.8 Å PaPPase:Mg_5IDP (maroon carbon, blue IDP, cyan Mg^{2+}) structure.

4.3.5.5. Comparisons with Previous Structures in Helical Conformations

There was a noticeable difference in helix 5 of the inner ring helices, wherein it formed a more restricted channel opening at the cytoplasmic side in PaPPase:Mg_5IDP compared with VrPPase:Mg_5IDP (Figure 4.17). Moreover, 3 of the outer ring helices (helices 2, 13 and 14) also appeared to be in different conformations to the VrPPase:Mg_5IDP structure. This could explain the difficulty in the refinement of these structures as the homology model was based on the placement of the helices in TmPPase:Mg_5IDP and

VrPPase:Mg_5IDP structures. These helices were repositioned during refinement to better fit the density and therefore represented a significant structural feature of PaPPase. The reasons for these differences are not known, but may be involved in the mechanism of potassium-independence in PaPPase. Higher resolution structures of PaPPase, or new structures of additional K^+-independent M-PPases would be needed to confirm this hypothesis.

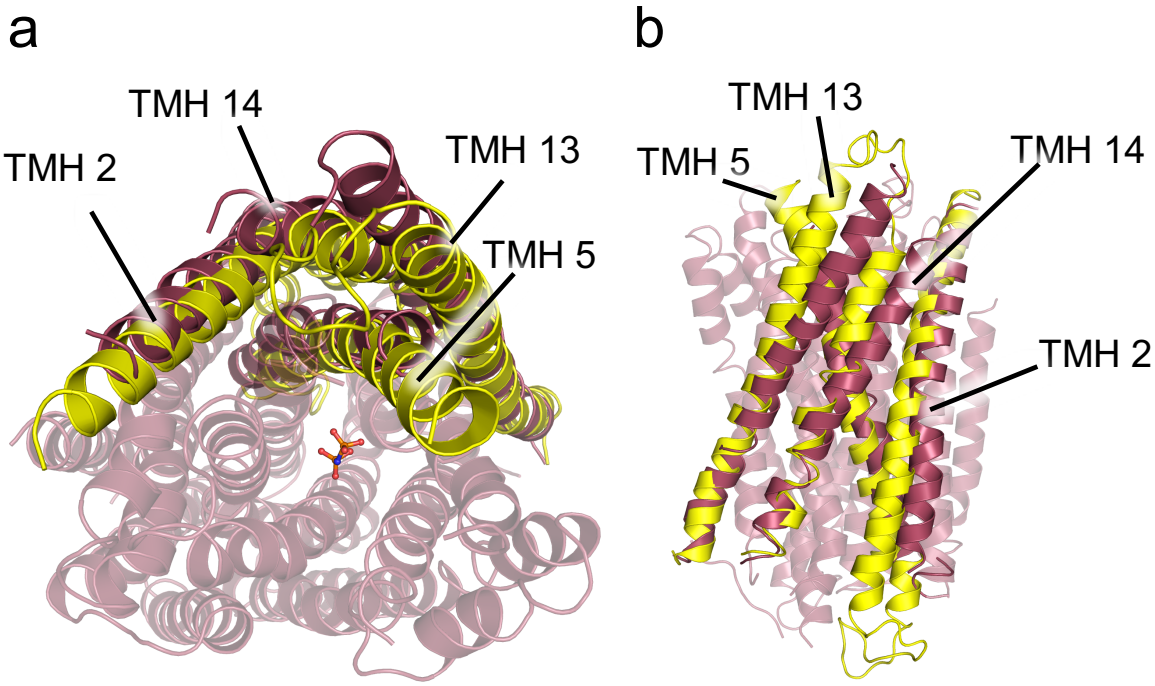


Figure 4.17: Comparison of Helix Conformations Between PaPPase and VrPPase
Comparisons between the helical conformations of 3.8 Å PaPPase:Mg_5IDP structure (maroon) and VrPPase:Mg_5IDP (yellow) showing (a) the top down view of the cytoplasmic side and (b) the side view of the proteins.

4.3.5.6. Comparisons with Previous Structures at the Ion Gate

As expected, there was no extraneous electron density at the ion gate of the PaPPase structure (Figure 4.18a), $2F_o$-F_c density was observed for $D^{6.50}$ and $S^{6.54}$ but there was no observable electron density for $N^{16.46}$, $K^{16.50}$, $E^{6.53}$ or $R^{5.50}$ and so the conformations of the side chains are based purely on the stereochemical properties of these residues (Figure 4.18). A comparison was made of residues at the ion gate region of PaPPase:Mg_5IDP and VrPPase:Mg_5IDP (Figure 4.18b). The position of helix 6 had moved by around 0.5 Å from the cytoplasmic side in PaPPase:Mg_5IDP but the overall position of the residues of the ion gate was very similar, with differences in resolution, and the resulting lack of electron density, likely the cause of the discrepancies.

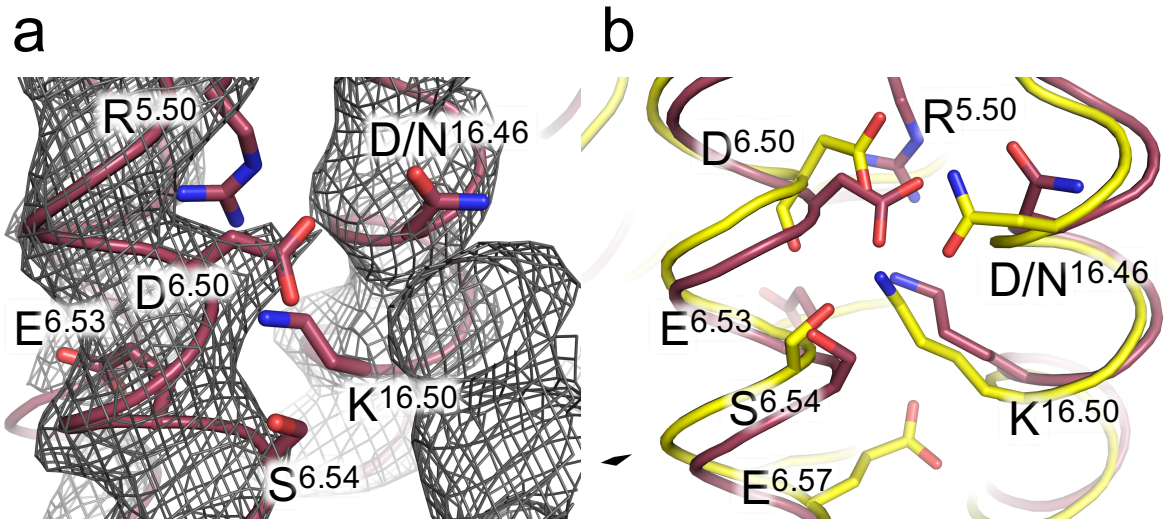


Figure 4.18: Ion Gate Overview of 3.8 Å PaPPase Structure.
Ion gate region of the 3.8 Å PaPPase:Mg_5IDP, showing (a) $2F_o$-F_c density (3 δ) following refinement and (b) a comparison between ion gate residues of PaPPase:Mg_5IDP (maroon) with VrPPase:Mg_5IDP (yellow).

4.3.6. Potassium-Independence of PaPPase

4.3.6.1. Structural Analysis and Identification of Potassium Independence Mechanisms

Previous mutational studies of K^+-dependent H^+-PPases (Belogurov & Lahti, 2002) had shown that the mutation of an alanine to a lysine in helix 12 could produce a K^+-independent mutant, with a modulatory effect seen with a glycine-to-threonine mutation one turn down the helix. Both the lysine and threonine are present in PaPPase at positions $K^{12.46}$ and $T^{12.49}$, respectively. There is electron density ($2F_o$-F_c) for both residues suggestive of side chain conformations (Figure 4.19a). A comparison was made of this region between the PaPPase:Mg_5IDP and VrPPase:Mg_5IDP structures (Figure 4.19b). From this it appears $K^{12.46}$ might be functionally replacing K^+. Although the NH_3 group does not directly occupy the same position as K^+, both interact with the IDP (Figure 4.19b). The limited resolution of the PaPPase:Mg_5IDP structure means that water molecules are not visible, but these coordinate K^+ in the VrPPase structure and it is highly likely that they would fulfil a similar role in the PaPPase structure.

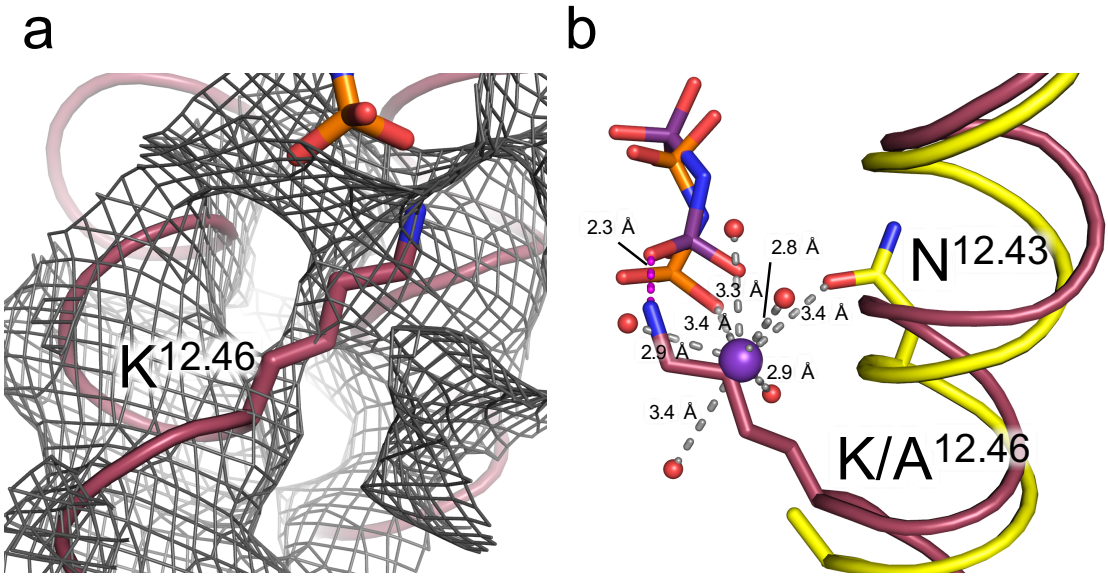


Figure 4.19: Overview of $K^{12.46}$ in 3.8 Å PaPPase Structure.
(a) $2F_o$-F_c density for $K^{12.46}$ of the 3.8 Å PaPPase structure following refinement. (b) comparison of the potassium binding site of VrPPase:Mg_5IDP (yellow carbon, orange IDP, purple K^+, red water) with the $K^{12.46}$ residue of the 3.8 Å PaPPase structure (maroon carbon, purple IDP). Mg^{2+} and water molecules not interacting with K^+ were removed for the sake of clarity.

The previous chapter (Section 3.6) showed that single mutations of either $K^{12.46}$A or $T^{12.49}$A in PaPPase were not sufficient to generate a K^+-dependent mutant (Figure 4.20). However, a dual mutant of both residues was K^+-dependent. Using the structural information obtained from the PaPPase:Mg_5IDP structure, it was possible to speculate on what was happening at K/A$^{12.46}$ and T/A$^{12.49}$ in each of the mutants. In the original study, the M-PPase of *Carboxydothermus hydrogenoformans* (ChPPase) had two alanine residues in place of the lysine and threonine found in most K^+-independent M-PPases. In PaPPase, $T^{12.49}$ is interacting with $D^{11.50}$ on helix 11. This residue therefore plays a role in stabilising PaPPase but is not present in any of the K^+-dependent M-PPases, including ChPPase. In K^+-dependent variants, the potassium ion is located in the active site, where it interacts with the electrophilic phosphate group of the substrate/inhibitor, as seen in the VrPPase crystal structures (PDB ID: 4a01). It was speculated that the NH_3 group of $K^{12.46}$, conserved in all K^+-dependent variants, is a substitute for K^+ (Belogurov & Lahti, 2002). Therefore, it was interesting to discover that the $K^{12.46}$A mutation alone did not appear to generate a K^+-dependent variant of PaPPase in this study.

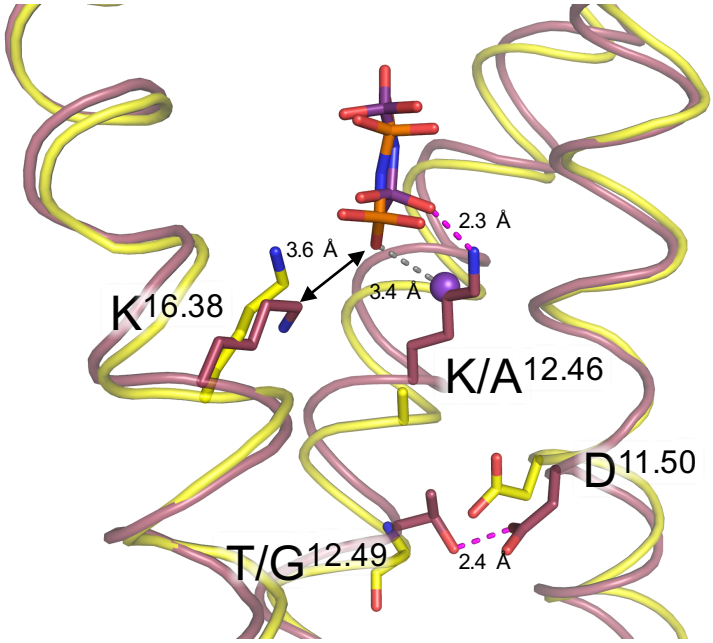


Figure 4.20: Overview of the Potassium-Independence Mechanism of PaPPase.
Positions of the two tested residues of K^+-independence ($K/A^{12.46}$ and $T/G^{12.49}$) and interactions made with neighbouring residues in PaPPase:Mg_5IDP (maroon carbon, purple IDP) and VrPPase:Mg_5IDP (yellow carbon, orange IDP, Purple K^+).

4.3.6.2. Comparison of PaPPase Structure with Previous Mutagenesis Results

Given the total conservation of $K^{12.46}$ in K^+-independent M-PPases and its absence in all K^+-dependent variants, it is clear this residue has a crucial role in K^+-stimulated selection. Previous studies on generating a K^+-independent variant from the K^+-dependent ChPPase found that in the absence of K^+, Wild-type and the $A^{12.49}T$ mutant of ChPPase showed no detectable proton pumping activity (Belogurov & Lahti, 2002). Increasing concentrations of K^+ from 0.1 to 100 mM increased the activity from 0 up to around 7 µmol min^{-1} mg^{-1}. Inversely, the $A^{12.46}K$ mutant had a constant activity of 4 µmol min^{-1} mg^{-1} across the same range of K^+, showing activity even in the complete absence of K^+. A later study of Na^+-PPases, all of which are K^+-dependent, showed hydrolytic activity of 20 µmol min^{-1} mg^{-1} in the presence of 50 mM K^+ and at least 1 mM Na^+ (Belogurov *et al.*, 2005). The activity was only 0.5 µmol min^{-1} mg^{-1} at K^+ concentrations of 0.004 mM when Na^+ concentrations were below 1 mM. However, activity increased with increasing concentrations of Na^+ between 1 and 100 mM, at which the activity was around 10 µmol min^{-1} mg^{-1}. Further study on the Na^+-PPase from *Methanosarcina mazei* (MmPPase) confirmed that K^+-dependent M-PPases could be activated by Na^+, even in the absence of K^+ (Malinen *et al.*, 2007).

4.3.6.3. Potential Role of Sodium Ions in Potassium Independence

Based on all the available data, Na^+ can substitute K^+ in K^+-stimulated PPase activity. The activity assay results for the PaPPase mutants are likely thrown off by the presence of Na^+ in the buffers (Figure 3.15). The structure of PaPPase confirms the presence of $K^{12.46}$ around the K^+-binding site seen in previous structures (Kellosalo *et al.*, 2012; Lin *et al.*, 2012) (Figure 4.19, Figure 4.20). Therefore, it is likely that $K^{12.46}$A has Na^+ occupying the K^+-binding site, compensating for the loss of the NH_3^+ group of the lysine. The additional interactions made between $T^{12.49}$ and $D^{11.50}$ provide a stabilisation mechanism that maintains the integrity of the K^+-binding site. This interaction maintains full activity in the wild-type and $K^{12.46}$A mutant of PaPPase, by allowing the lysine or Na^+ to interact with the substrate. The loss of this interaction in the $T^{12.49}$A mutant reduces the stability of the K^+-binding site. The lysine therefore cannot interact as well with the substrate, so the overall activity is reduced, but K^+ or Na^+ was still not required for activity. The dual mutant lacks the NH_3^+ substitution for K^+ but also lacks the stabilising effects of the $T^{12.49}$-$D^{11.50}$ interactions. The overall activity is therefore significantly reduced compared with wild-type. It is possible that the K^+ in the K^+-positive sample buffer of the activity assay provides the increased monovalent cation concentration required to recover activity in the dual mutant, as seen in the previous studies of K^+-dependence (Belogurov *et al.*, 2005; Malinen *et al.*, 2007). It could mean the K^+-dependent effect seen here is not a true effect and is related to the increasing monovalent cation concentration. Future experiments on the PaPPase mutants, without the presence of any monovalent cation, will be needed to fully confirm the role of the $K^{12.46}$A mutation in the generation of K^+-dependence in PaPPase and to confirm the original findings by Belogurov and Lahti (2002) that this residue alone can account for K^+-stimulated selectivity.

4.3.6.4. Alternative Mechanisms to Explain Potassium Independence

The effects of Na^+ substitution for K^+ were only studied in Na^+-PPases and not H^+-PPases. Since the mechanism of cation dependence between Na^+ and H^+ has not been identified, it is also possible that Na^+ cannot substitute K^+ in H^+-PPases. Therefore, hydrolytic activity retained in the $K^{12.46}$A mutant may represent a genuine effect of K^+-independence. It is possible that the stability of the $T^{12.49}$-$D^{11.50}$ interactions allows the protein to compensate for the loss of K^+ and $K^{12.46}$ by the substitution of this charge with

other nearby positively-charged residues, such as $K^{16.38}$, located 3.6 Å from the substrate (Figure 4.20). Conformational changes in the side-chain of this lysine residue may provide a compensatory mechanism that allows the protein to remain active even in the absence of K^+. The conformational changes may not be possible in the dual mutant with the loss of the stabilising effects of the $T^{12.49}$-$D^{11.50}$ interaction. The loss of stability would affect the overall activity of this mutant, but would allow K^+ to bind still at the expected K^+-binding site, recovering the activity. This means the K^+-dependent effect seen for the dual-mutant could be a genuine effect.

The results above show that the structure of wild-type PaPPase alone cannot explain the K^+-dependence mechanism of M-PPases. Future work will need to be carried out on the mutants generated in this study, ensuring a complete absence of Na+ from the buffers. This could confirm if Na^+ can substitute K^+ in the $K^{12.46}$A mutant and potentially confirm the previous findings that this lysine is the sole determinant of K^+-dependence in M-PPases (Belogurov & Lahti, 2002). Crystal structures of the two PaPPase mutants could also provide valuable information on whether the compensatory mechanism was true or could provide additional confirmation that Na^+ was substituting K^+.

4.4. Summary of M-PPase Structural Studies

My aim was to carry out structural studies on TmPPase and PaPPase, generate structures in multiple catalytic states and use this information to understand how the protein functions on a mechanistic level, related to ion pumping, hydrolysis and K^+-dependence. In this way, I could compare, contrast and combine the new structures with the previously solved structures to generate a complete model of the catalytic cycle to allow for more in-depth analysis.

I solved two new structures of TmPPase and structures of a new M-PPase, the PaPPase. The new structures of TmPPase were in two new states compared with the previously solved structures of TmPPase (Kellosalo *et al.*, 2012). This means there are now four complete states including the resting state, substrate-analogue-bound state (IDP), 2-product-bound state (P_{i2}) and the single product-analogue bound state (WO_4^{3+}). The structures solved by our collaborators, both previously (Lin *et al.*, 2012) and in line for this project and publication (K. M. Li *et al.*, 2016) were of VrPPase in the substrate-analogue bound state (IDP) and now in the single product-bound form (P_i).

The TmPPase:Mg_5IDP structure may be a subtly distinct sub-state of the substrate-bound state of the protein. The hydrolytic centre of both TmPPase:Mg_5IDP and VrPPase:Mg_5IDP, were very similar, including the position and coordination of the Mg_5IDP complex and the overall position of the helices. There was a noted "downwards" motion of helix 12 by 2 Å away from the cytoplasmic side when comparing the resting state structure to the IDP-bound structures. Indeed the only significant different between TmPPase:Mg_5IDP and VrPPase:Mg_5IDP was at the ion gate, where a sodium ion was bound in the TmPPase structure.

The TmPPase:$MgWO_4$ structure shares a high degree of similarity to the previous TmPPase resting state structure (Kellosalo *et al.*, 2012). The unique feature is the presence of the $MgWO_4$ complex in the active site, which occupies the same position as the single-product bound structure (K. M. Li *et al.*, 2016). Overall it is clear that both the TmPPase:$MgWO_4$ and VrPPase:Mg_5P_i represent the same state, whereas TmPPase:Mg_5IDP and VrPPase:Mg_5IDP clearly represent two distinct sub-states of the overall substrate-bound state. These structures, along with the previously solved structures (Kellosalo *et al.*, 2012; Lin *et al.*, 2012) will be analysed in much more detail in chapter 5.

Initial crystallisation attempts of PaPPase produced crystals that diffracted to around 20 Å. Through optimisation of conditions, attempts using different types and concentrations of detergents and various forms of dehydration experiments, the resolution was steadily improved down to 3.8 Å. As the resolution improved more structural and mechanistic information became available. Helices 13 and 14 showed a noticeable change in position compared with the previous IDP-bound structures of TmPPase and VrPPase. Although the reasons for this are unclear, they may have a role in the K^+-independence mechanism of PaPPase.

Despite the presence of positive density (F_o-F_c) for the Mg_5IDP complex, the 5-6 loop, which packs above this in the TmPPase:Mg_5IDP and VrPPase:Mg_5IDP structures, was only partially visible in the PaPPase:Mg_5IDP structure. Indeed, the entire cytoplasmic side of the protein appears to form less contacts than the previous structures, accounting for the lack of electron density for these regions as they are disordered. Improving the

crystal packing and increasing the resolution of the structure will be necessary to gain more detailed mechanistic information about PaPPase. However, the current structure provides the first major overview of this protein including the key catalytic residues of the active site and some of the residues around the ion gate. This will allow for a detailed comparison with the currently available TmPPase and VrPPase structures during the generation of the catalytic cycle model in chapter 5.

Chapter 5: Characterising the Membrane Pyrophosphatase Catalytic Cycle

5.1. Introduction to the M-PPase Catalytic Cycle

The novel crystal structures of TmPPase and VrPPase described in chapter 4 have provided new insights into the different catalytic states of M-PPases. There now exists 6 structures of TmPPase and VrPPase with most of these in different catalytic states. By comparing differences in each of these static snapshots, a model of the complete M-PPase catalytic cycle could be generated to provide useful information on how these proteins function. Three previously solved structures exist; TmPPase in the resting state and product-bound (Mg_5P_{i2}) state and VrPPase in the substrate-bound state (IDP). I was able to solve two new structures of TmPPase (Chapter 4), in the substrate-bound state (IDP) and product-bound state ($MgWO_4$). An additional structure of the VrPPase, in the immediate product-bound state (Mg_5P_{i2}), was simultaneously solved by our collaborators from the Sun lab (Taiwan). Through the work described in this chapter, I analysed these structures and compared the residue side chains and helical positions, looking at the helical geometry and hydrogen bonding patterns between structures of the various catalytic states. Previous catalytic schemes had been proposed for M-PPases and the new structural information was used to try and better understand the cycle.

I also carried out tryptophan fluorescence and proton translocation studies to try and understand the states that are difficult to resolve using static crystallographic snapshots, such as the high-energy transition state. This state involves the opening of the ion gate with cation release from the channel, but currently no structural or mechanistic information exists on how this occurs. Crystal structures show a state containing the substrate-analogue with the proton/sodium ion bound, but the next structure shows product bound but no ion/proton. This means the order of events during this transition state was not known, nor was the mechanism of gate opening or ion/proton release. Measuring the change in tryptophan fluorescence upon substrate binding would show potential changes in conformational states. If the measured change was large enough, this would confirm the possibility of using stopped-flow fluorescence experiments to measure changes in conformation throughout the entire catalytic cycle. Measuring both the change in fluorescence upon substrate binding and ion pumping provided some time-resolved data to piece together what was happening during the catalytic cycle and, more importantly, what was happening during the transition state.

5.2. ANS Fluorescence Experiments to Study Conformational Changes in M-PPases

To better understand the catalytic cycle, I sought more time-resolved methods to observe various conformational changes occurring over time and try to link these to different states of the catalytic cycle. The hope was to find states that corresponded to the crystal structures and potentially identify new ones that could not be found by standard crystallographic techniques, such as the high-energy transition state. I studied the tryptophan fluorescence of the M-PPases, as this technique can be used to show changes in overall protein conformations. Initial tests were done using a standard fluorimeter and titrations of the substrate-analogue IDP as this was known from the crystal structures to cause the formation of the closed conformational state.

The transition from open to closed state was predicted to be the largest conformational change and was thus tested as a proof-of-concept. If a sufficient change in fluorescent signal was observed, it might be possible to carry out stopped-flow experiments to try and observe conformational changes occurring over the millisecond timescale using real-time fluorescence measurements. To try and boost the signal generated by tryptophan fluorescence, 8-anilino 1-naptholenesulphonate (ANS) was used, which fluoresces when activated by the emission peak of tryptophan (maximal 370 nm) and has a maximal emission peak at 450 nm. This compound binds to any hydrophobic region of the protein and significantly amplifies the fluorescent signal as there are orders of magnitude more ANS present per dimer than tryptophan residues in the protein. However, since it is only activated by tryptophan fluorescence emission, if there is no signal present to start with, it cannot amplify it and so no signal will be seen. Any change in fluorescence signal will either be caused by changing conformations of tryptophan residues or due to additional displacement/binding of ANS as the overall conformation of the protein changes. This means there are several causes for a change in fluorescent signal during the experiments, but all of them imply a degree of conformational change, which was the rationale for the experiments.

5.2.1. Fluorescence Studies on TmPPase

I carried out initial fluorescence experiments on DDM-purified TmPPase (Chapter 3), starting with titrations of the inhibitor and substrate analogue imidodiphosphate (IDP).

Experiments were carried out using biologically relevant concentrations of IDP (0-500 µM) (50 x the expected K_d) and magnesium (5 mM) (Figure 5.1, Figure 5.2), during which no change in the fluorescence signal was observed. This was backed up by the binding curve (Figure 5.2), which also shows no change in fluorescence quench with increasing concentration of IDP. It was therefore not possible to calculate K_d values for IDP binding to TmPPase using tryptophan fluorescence techniques. Structural analysis of the four TmPPase structures revealed seven tryptophan residues per protomer. Between the open-state (PDB ID: 4AV3) and the closed-state (TmPPase:Mg_5IDP, current study), none of these were in positions undergoing significant conformational changes between these two states. Therefore, it stands to reason that no fluorescent signal would be detected during these experiments.

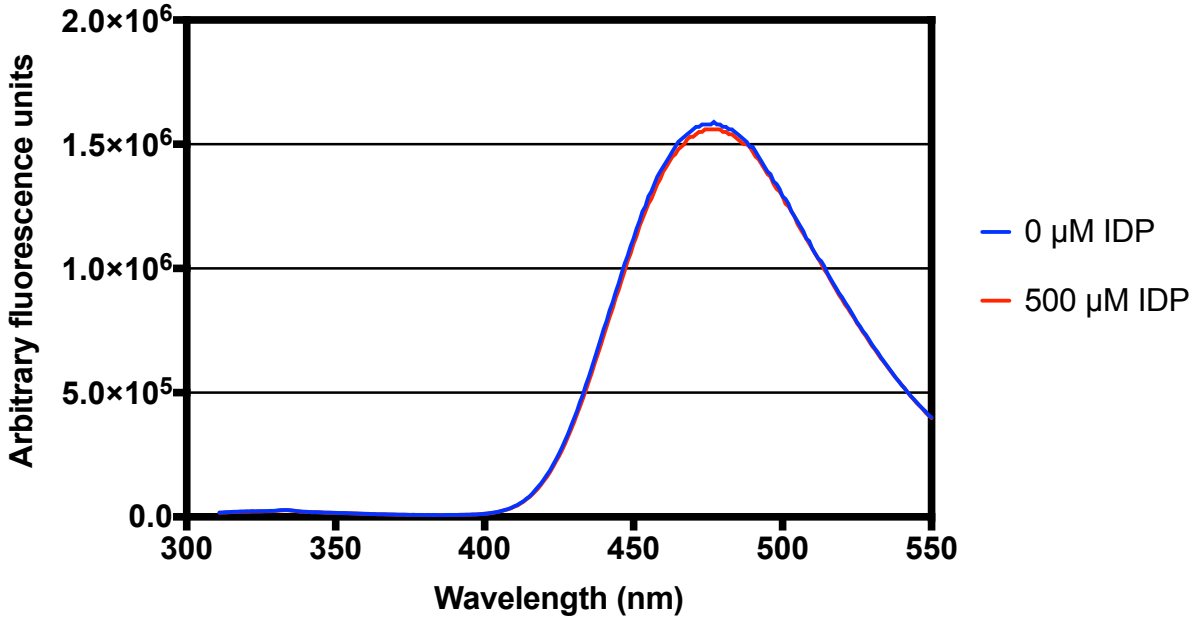


Figure 5.1: Tryptophan and ANS Fluorescence of TmPPase
Imidodiphosphate titrations at biologically relevant levels during TmPPase fluorescence experiments showing emission spectra of ANS following excitation of tryptophan residues at 295 nm.

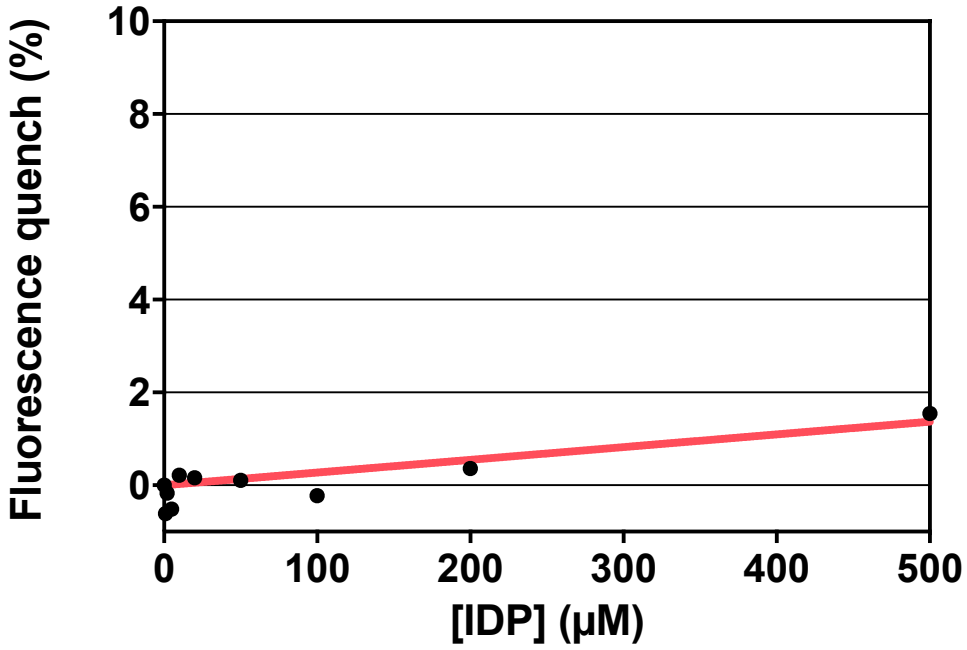


Figure 5.2: Effect of IDP concentration on ANS Fluorescence in TmPPase.
Percentage of fluorescence quenching of TmPPase against biologically relevant concentrations of imidodiphosphate in the prescence of 5 mM $MgCl_2$,

5.2.2. Fluorescence Studies on VrPPase and PaPPase

An analysis of the amino acid positions in VrPPase and PaPPase showed that tryptophan residues were present in regions of the protein that undergo significant conformational changes throughout the catalytic cycle (based on comparisons of expected residue positions between the TmPPase resting-state and IDP-bound states). Moreover, in VrPPase one of these tryptophan residues was located in the 5-6 loop, known to undergo extensive conformational changes as it forms the main mechanism through which the protein seals the channel upon substrate binding.

I carried out tryptophan fluorescence experiments on samples of purified PaPPase (Chapter 3) and VrPPase (provided by the Sun lab, Taiwan). In PaPPase, the fluorescence signal changed by around 4-5 % when IDP was titrated over a biologically relevant range (0-500 μM) (Figure 5.3). The K_d had not been determined previously for PaPPase, but was found to be 5.7 μM, around the observed level for other M-PPases (~12 μM) (Baykov *et al.*, 1993). Since this change in fluorescence signal is linked to IDP concentration this likely represents conformational changes in the protein. However, the percentage change in fluorescence signal is too weak for use in future stopped-flow experiments, as the

signal from smaller or more subtle changes in conformation would be drowned out by background noise.

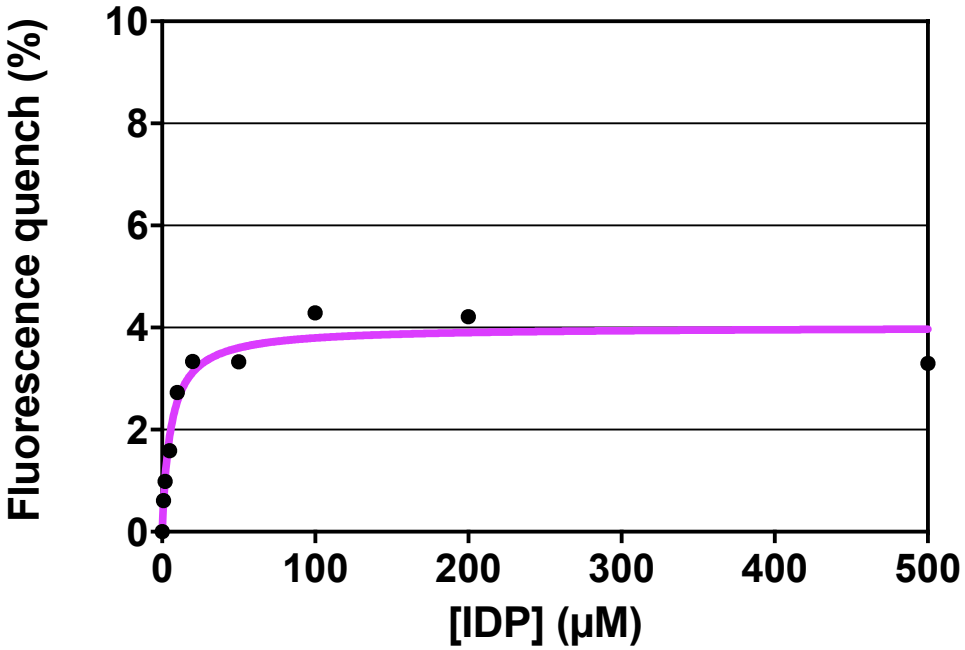


Figure 5.3: Effect of IDP concentration on ANS Fluorescence in PaPPase
Percentage of fluorescence quenching of PaPPase against biologically relevant concentrations of imidodiphosphate in the prescence of 5 mM $MgCl_2$.

The results from the VrPPase experiments showed a change in fluorescence signal linked to the titration of IDP (Figure 5.4). The maximum change in fluorescence signal was determined to be around 8 %, with a K_d value of 6.9 μM, around the same level previously identified for *R. rubrum* M-PPase (Baykov *et al.*, 1993).

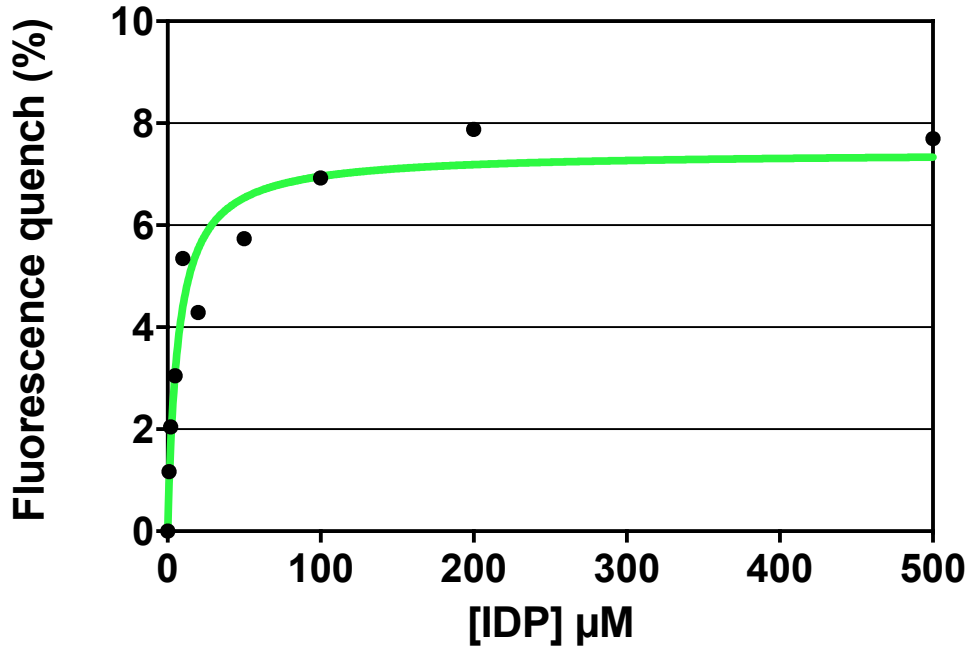


Figure 5.4: Effect of IDP concentration on ANS Fluorescence in VrPPase
Percentage of fluorescence quenching of VrPPase against biologically relevant concentrations of imidodiphosphate in the prescence of 5 mM $MgCl_2$.

5.2.3. Summary of Tryptophan Fluorescence Results

In M-PPases, the largest predicted conformational change was the initial substrate binding and closure of loop 5-6. The initial fluorescence experiments were carried out to see if this conformational change could be detected through measurements of tryptophan fluorescence upon binding of the inhibitor, IDP. The maximal change in fluorescence observed was on the borderline required to gain a clear signal above background during stopped-flow experiments. TmPPase and PaPPase would not be suitable for future fluorescence experiments due to the weak or absent change in fluorescence signal, but VrPPase showed a change of 8 %, which might allow for stopped-flow fluorescence analysis. Due to time constraints, the stopped-flow experiments could not be carried out as part of this project. However, the VrPPase sample represents the best opportunity of the three proteins tested, to see conformational changes occurring using the stopped-flow fluorimeter. These future experiments would involve titrations of PP_i to see if multiple conformations could be detected. If so, it would then be possible to start assigning these to known catalytic states using kinetic information to generate a detailed model of the catalytic cycle, supplementing currently available structural information.

5.3. Nanion SURFE^{2}R Experiments to Study Proton Pumping in M-PPases

The SURFE^{2}R is a relatively new technique developed by Nanion Technologies (Germany) and is a modified form of solid-support membrane (SSM) based electrophysiology. It is designed to study charged transport by proteins reconstituted into liposomes. The technique works by adhering proteoliposomes to a membrane layer affixed to a gold sensor chip. The experiments typically occur over a 3 second time period. The buffer can be rapidly switched from a control buffer to an activating buffer, containing the substrate or ligand of interest. The activating buffer is added at the 1 second time point, there is a short (~0.1 second) delay during which the buffer exchange occurs, substrate or ligand binds and pumping of charged molecules starts to occur into the liposomes. This charged pumping can be measured directly as current, an overview of the technique is given in Figure 2.4.

These experiments were carried out on the proton-pumping VrPPase protein (purified protein provided by the Sun lab, Taiwan). This was reconstituted into liposomes following previously described protocols used during earlier patch-clamp studies on VrPPase (Nakanishi *et al.*, 2003), where the PP_i-induced ion pumping signal was observed, that was inhibited by IDP. It was hoped that the SURFE^{2}R technique would be a less complicated alternative to study ion pumping in M-PPases, allowing us to compare kinetic data under different buffer conditions or using different mutants of important ion pumping residues. Initial experiments involved testing VrPPase on the system to see if a proton-specific signal could be observed.

5.3.1. Proton Pumping Studies of VrPPase

5.3.1.1. Initial Characterisation of the Proton Pumping Signal in VrPPase

Initial experiments were performed to characterise VrPPase on the SURFE^{2}R, using the PP_i substrate to see if a signal could be generated. Results were calculated as a change in current measured as the subtraction of the maximal peak, obtained from proton pumping, from the baseline prior to activating buffer addition. K_4PP_i was used in the first instance and a change of current of 3.13 nA (+/- 0.39 nA) was detected when it was added at 50 µM to the activating buffer (Figure 5.5). Variants of the substrate, such as Na_4PP_i, yielded a change in signal of 2.87 nA (+/- 0.4 nA), which showed no statistically significant

difference to K_4PP_i using an unpaired T-test (Figure 5.12). This showed that the accompanying cation was not important for activity (Figure 5.5)

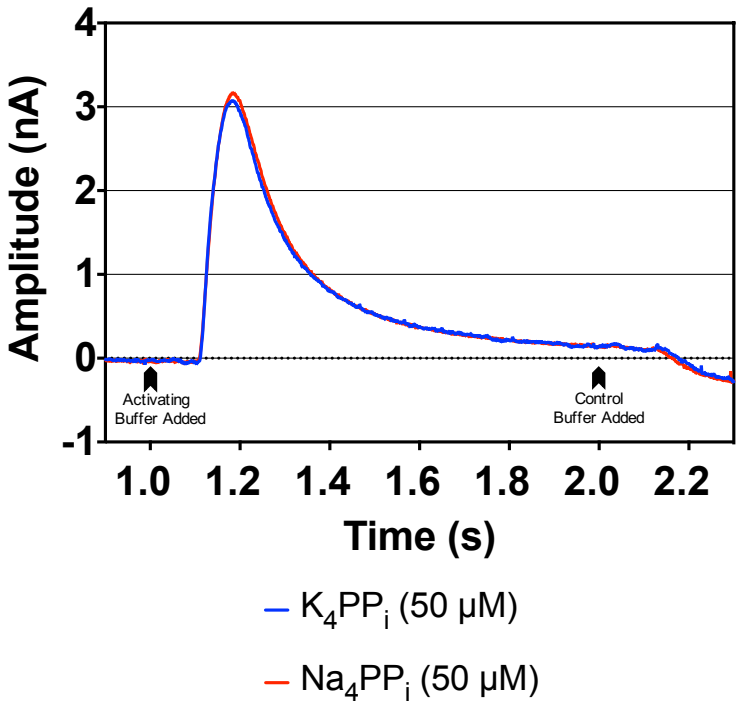


Figure 5.5: Effect of Substrate Type on Proton Pumping in VrPPase Proteoliposomes.
Initial SURFE^{2}R results showing current over time following addition of the activating buffer at 1 second, containing either K_4PP_i or Na_4PP_i, and return to control buffer at the 2 second time point.

In addition to the experiments with PP_i in the activating buffer, control experiments were carried out with P_i-containing compounds or water in the activating buffer. The addition of equal volumes of water to the activating buffer yielded no signal, implying the signal in the main experiment was due to a PP_i-induced effect on the protein and not due to any kind of signal generated by the transition of charged molecules across the sensor during the buffer exchange. The use of P_i in the activating buffer generated a much smaller change in current of 0.067 nA (+/- 0.03 nA), significantly below that seen for even the lowest concentrations of PP_i tested (Figure 5.6, Figure 5.12). This represents the background signal, accounting for any potential current generated via the buffer exchange step (movement of charged molecules). This signal does not represent proton pumping as this does not occur during binding of P_i to the active site. The crystal structures show that P_i can induce a closed conformational state, making it a good negative control for these experiments (Kellosalo *et al.*, 2012). Moreover, the phosphate was added at 4 x the

maximal concentration of PP_i used to ensure that this background effect was accounted for.

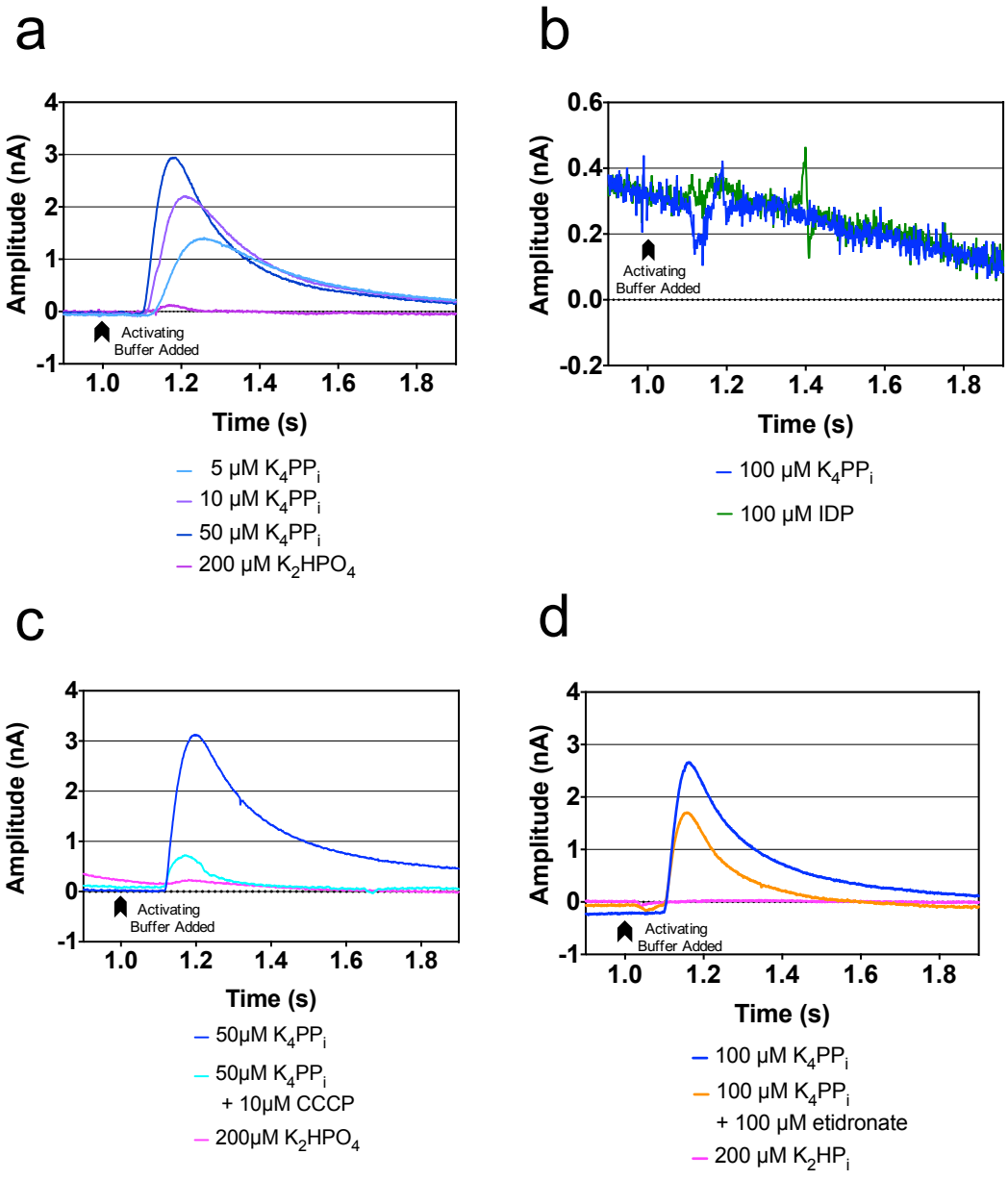


Figure 5.6: Confirmation of Proton Pumping Signal in VrPPase Proteoliposomes.
SURFE^{2}R results showing current over time following the addition of (a) increasing concentrations of K_4PP_i substrate to VrPPase proteoliposome samples, (b) substrate and inhibitor to empty liposome samples, (c) K_4PP_i substrate with and without the protonophore, CCCP and (d) K_4PP_i with and without the inhibitor, etidronate. All compounds were added to the activating buffer and injected into the system at 1 second, control buffer was returned to the system at the 2 second timepoint. Phosphate control samples substituted substrate in the activating buffer of (a), (b) and (c).

Additional control experiments were also performed on sensors containing empty liposomes. Experiments were run with PP_i, P_i and water and no signal was observed, even with 100 µM PP_i (Figure 5.6b). This shows that the signal seen in the main

experiments was not only a PP_i-specific signal, but also a protein-specific signal, since the PP_i-only experiments did not yield any current in the absence of protein. The next step was to verify that the signal produced was due to proton pumping and not any other PP_i-induced effect on the protein. To achieve this, several ionophores were used. These compounds dissipate the membrane potential by allowing specific ions to pass across the membrane. The initial compound used was gramicidin, which completely dissipates the membrane potential by allowing the passage of different cations across the membrane. Experiments were also repeated with a proton-specific ionophore, Carbonyl cyanide m-chlorophenyl hydrazine (CCCP), in place of gramicidin.

The ionophores were added to a final concentration of 10 μM to all buffers during the experiment and the resulting signal caused by the addition of PPi was significantly reduced from 3.13 nA (+/- 0.38 nA) to 0.65 nA (+/- 0.01 nA) (Figure 5.12), but not abolished (Figure 5.6c). This might be caused by the rate of proton pumping exceeding the rate of proton leakage from the liposomes, generating a signal despite the presence of the protonophore in the membrane. These experiments show that the signal is indeed due to cation pumping and not due to the effects of charged residues moving during conformational changes or due to signal generated by the effect of adding charged molecules during buffer exchange. These results show conclusively that the observed signal is generated by proton pumping and not any other indirect effect of PPi binding to the protein.

In SURFE^{2}R experiments, small differences in the ionic balance of the two buffers can generate background signal that would drown out weaker signals. Therefore, my early experiments on VrPPase sought to test different buffering conditions. These included the addition of K_2HP_i to the control buffer at concentrations that would ionically balance the activating buffer (e.g. 100 mM K_2HP_i to 50 mM K_4PP_i). The addition of K_2HP_i in the control buffer leads to a weaker signal of around 2.5 nA instead of 3.0 nA seen when no buffering component is added (Figure 5.7). One possible explanation is that the lack of a buffering compound leads to a 0.5 nA increase in signal generated solely by the binding of the charged PP_i molecule to the active site of the protein. It could also represent an inhibitory effect of the phosphate on the activity of the protein, as phosphate would compete with pyrophosphate for binding in the active site. Results from the SURFE^{2}R experiments do not represent specific activity of the protein as the concentration of

correctly orientated protein cannot be quantified. Therefore, it did not matter which method was used for future experiments, as long as it was kept consistent.

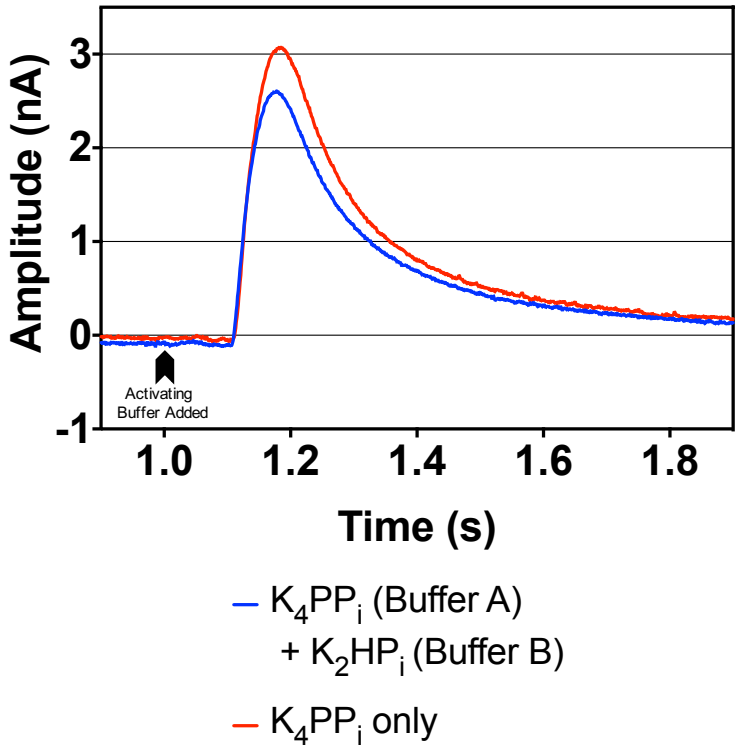


Figure 5.7: Ionic Balance Test of SURFE^{2}R Buffers.
ionic balancing experiments for the SURFE^{2}R results showing current over time following the addition of K_4PP_i in the activating buffer and samples with and without K_2HP_i in the non-activating buffer. Non-activating buffer was added at the 0 second timepoint, followed by activating buffer after 1 second and back to non-activating buffer at the 2 second timepoint.

5.3.1.2. Proton Pumping Studies using Inhibitors of M-PPases

The SURFE^{2}R system had been tested and there was strong evidence to suggest that proton pumping could be observed via the addition of the PP_i substrate to the activating buffer. Follow-up experiments were carried out using various pyrophosphate-analogues and inhibitors of the enzyme in place of the PP_i substrate in the activating buffer to see what effect these would have on proton-pumping. The first inhibitor tested was imidodiphosphate (IDP), which had been used for the crystallisation experiments. A change in signal of 0.29 nA (+/- 0.06 nA) was observed following the addition of 50 µM IDP (Figure 5.8), which represented a statistically significant difference to both the K_4PP_i (P = and K_2HPO_4 samples (Figure 5.12). This was a repeatable signal that was observed across multiple sensor chips. Background signals generated by the substitution of IDP with K_2HP_i generated a signal of only 0.067 nA (+/- 0.03 nA) (Figure 5.8). The addition

of gramicidin or CCCP to the IDP-containing buffer reduced the signal to 0.07 nA (+/- 0.03 nA), which was around the background level (Figure 5.8, Figure 5.10, Figure 5.13). Addition of etidronate alongside IDP in the activating buffer led to a reduction in the signal generated (Figure 5.9). Moreover, no signal was generated when the same concentration of IDP was used in the activation buffer on the no-protein control (Figure 5.6b). This confirms that the signal generated was also due to proton-pumping and not any indirect effect such as buffer exchange or the movement of charged residues of the protein during conformational changes. Since IDP is non-hydrolysable, this provides preliminary evidence to suggest that proton pumping is occurring prior to hydrolysis in the catalytic cycle. In order to test this hypothesis, additional inhibitors were tested (Table 5.1).

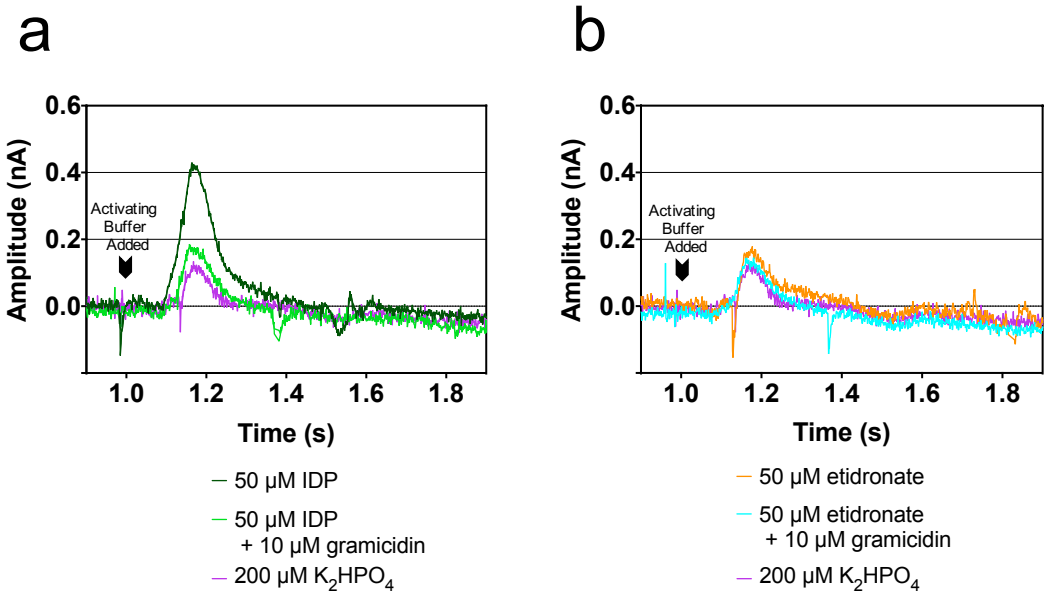
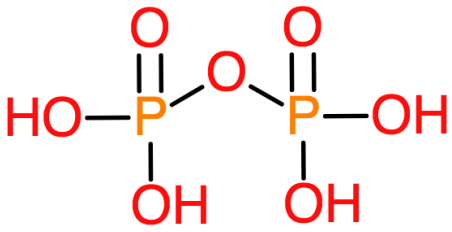
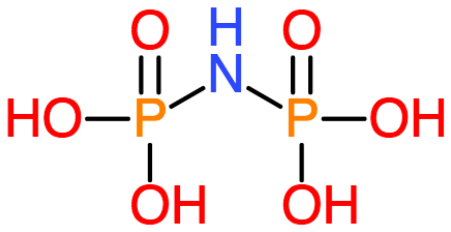
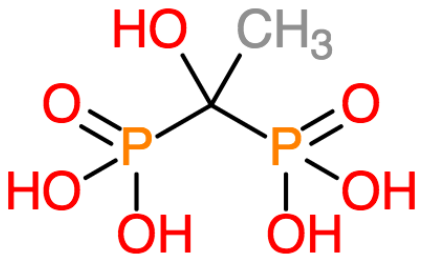
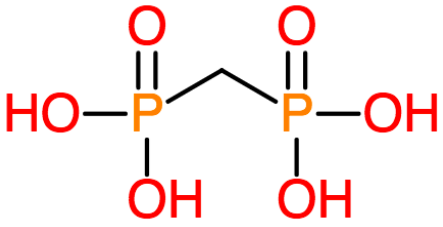


Figure 5.8: Inhibitor Studies with IDP and etidronate on Proton Pumping in VrPPase Proteoliposomes. *SURFE^{2}R results from inhibitor experiments showing current over time following the addition of (a) IDP with and without gramicidin, (b) etidronate with and without gramicidin, all compounds were added to the activating buffer and injected into the system at 1 second.*

Table 5.1: Pyrophosphate and Analogues Used during SURFE2R Experiments.

Name	Formula	Structure
Pyrophosphate (PP_i)	$O_3P\text{-}O\text{-}PO_3$	
Imidodiphosphate (IDP)	$O_3P\text{-}NH\text{-}PO_3$	
1-hydroxyethane 1,1-diphosphonic acid (etidronate)	$O_3P\text{-}C(CH_3)(OH)\text{-}PO_3$	
Methylene diphosphonate (MEDP)	$O_3P\text{-}CH_2\text{-}PO_3$	

The inhibitor 1-hydroxyethane 1,1-diphosphonic acid (HEDP), also known as etidronate (Table 5.1), was also tested in the activating buffer. This inhibitor contains a methyl and hydroxyl group off the central atom joining the two phosphates. During the experiments, no signal was seen above background, in the presence or absence of ionophores (Figure 5.8). The compound was modelled into the IDP-containing crystal structure of VrPPase in place of the IDP (Figure 5.11). The methyl and hydroxyl groups in etidronate would form steric clashes with $D^{15.61}$ and one of the magnesium ions required for the binding of the substrate. In this model, it would not be possible for the 5-6 loop to close correctly over the active site. If the loop cannot close then the enzyme would be completely

inactive, as the catalytic cycle would not be able to proceed in order to prevent the backflow of ions.

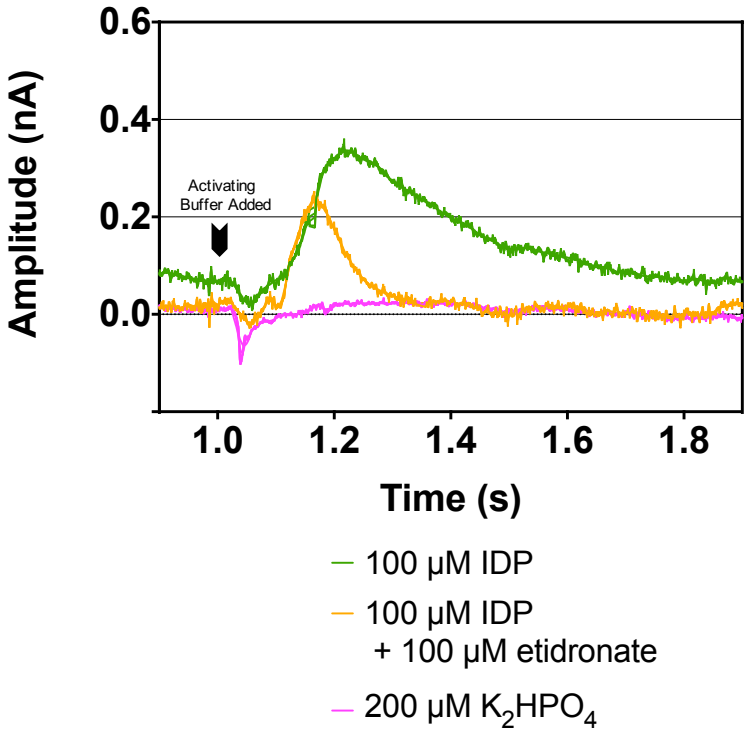


Figure 5.9: Effect of Etidronate on IDP-Induced Proton Pumping in VrPPase Proteoliposomes. *SURFE^{2}R results from inhibitor experiments showing current over time following the addition of IDP with and without etidronate. All compounds were added to the activating buffer and injected into the system at 1 second and returned to the control buffer at the 2 second timepoint.*

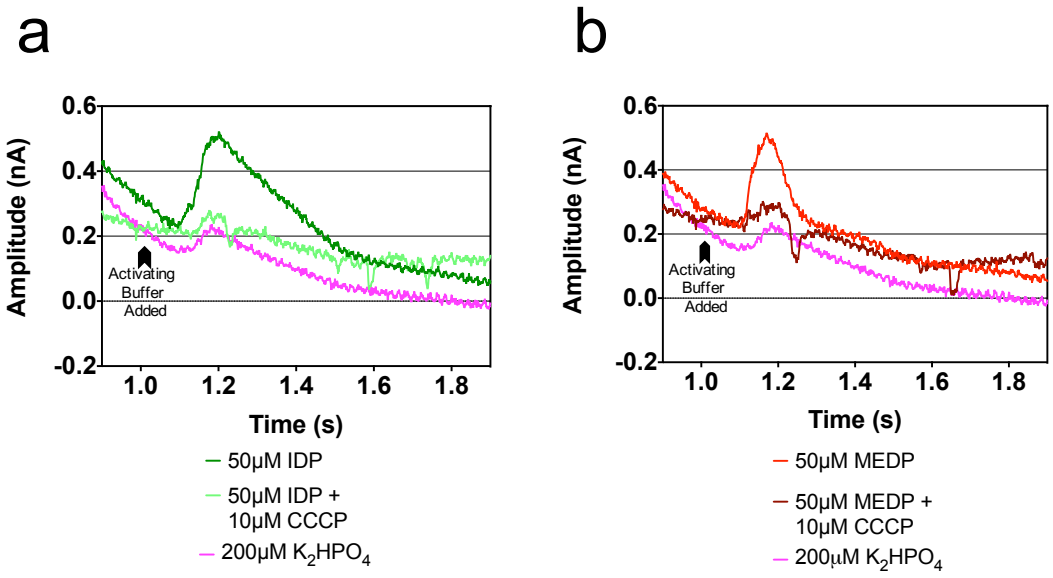


Figure 5.10: Effect of CCCP on Inhibitor-Induced Proton Pumping in VrPPase Proteoliposomes. *SURFE^{2}R results from inhibitor experiments showing current over time following the addition of IDP with and without the protonophore, CCCP and (e) MEDP with and without the protonophore, CCCP. All compounds were added to the activating buffer and injected into the system at 1 second, control buffer was returned at the 2 second timepoint..*

Methylene diphosphonate (MEDP) is another inhibitor of M-PPases that is structurally very similar to IDP and PP_i, and contained a carbon atom linking the two phosphate molecules. This inhibitor was also tested in place of PP_i in the activating buffer of the $SURFE^2R$ experiments on VrPPase. This compound also produced a change in signal of 0.24 nA (+/- 0.06 nA), which is similar to the IDP-induced proton pumping signal (Figure 5.13). signal when added as part of the activating buffer and again this signal was dissipated after adding CCCP or gramicidin (Figure 5.10). Since IDP and MEDP are structurally similar to PP_i (Table 5.1), they can still bind correctly at the active site (as seen in the crystal structure for IDP). This means loop 5-6 can still pack in over the active site to seal the channel, allowing the enzyme to proceed through the catalytic cycle up until the point of hydrolysis, at which point it will stall. These results therefore provide the first biochemical data to help in updating the catalytic cycle model, specifically providing information of the order of events during the catalytic cycle. The evidence of proton pumping in the presence of the non-hydrolysable inhibitor implies that proton-pumping is the first mechanism to occur, followed by hydrolysis of the substrate. The specific mechanism that couples ion pumping to hydrolysis cannot be ascertained from this information, but the information was invaluable in updating the model of the catalytic cycle, discussed in Chapter 6.

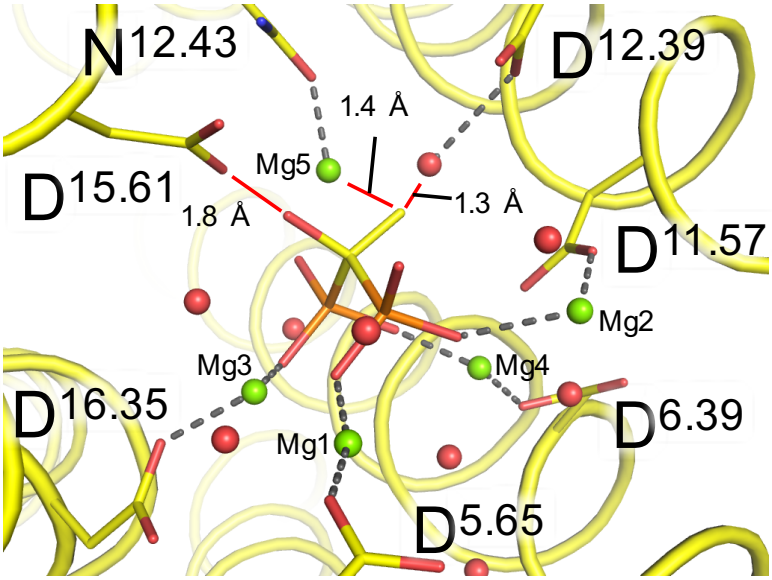


Figure 5.11: Model of Etidronate in the Hydrolytic Centre.
Modelled binding of the M-PPase inhibitor, etidronate in the active site of VrPPase. Inhibitor was modelled in place of the IDP molecule from the VrPPase:Mg_3IDP crystal structure (PDB ID: 4A01). Carbon atoms are shown in yellow, magnesium ions in green, water molecules in red and phosphorous in orange. Figure reused with permission (Supplementary figure 8, (K. M. Li et al., 2016))

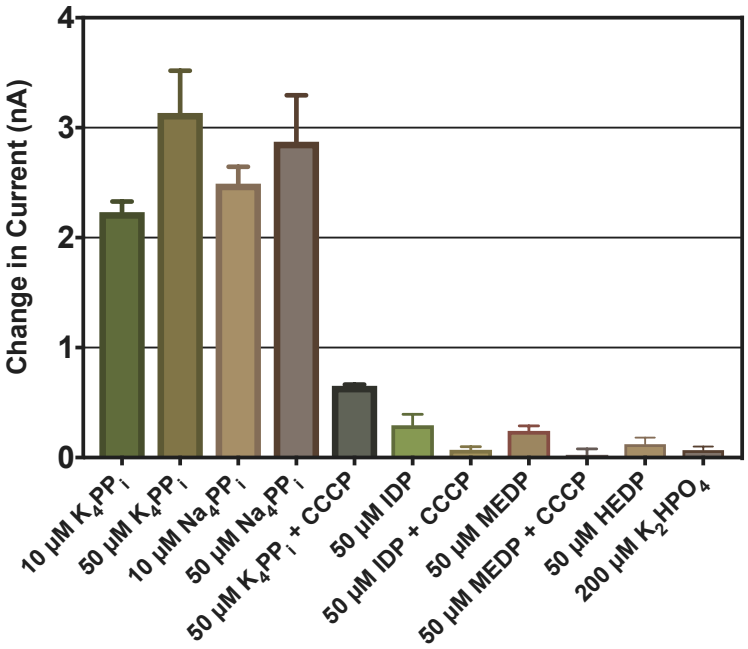


Figure 5.12: Effect of Different Compounds on Proton Pumping by VrPPase.
Effect of the substrate and various inhibitors on the proton pumping signal generated in VrPPase following their respective addition to the activating buffer at the 1 second time point of SURFE2R experiments. The change in current is determined by subtracting the maximum peak value obtained from proton pumping measurements from the baseline prior to addition of the activating buffer. All results are from multiple replicates carried out on multiple sensor chips and error bars denote standard deviation.

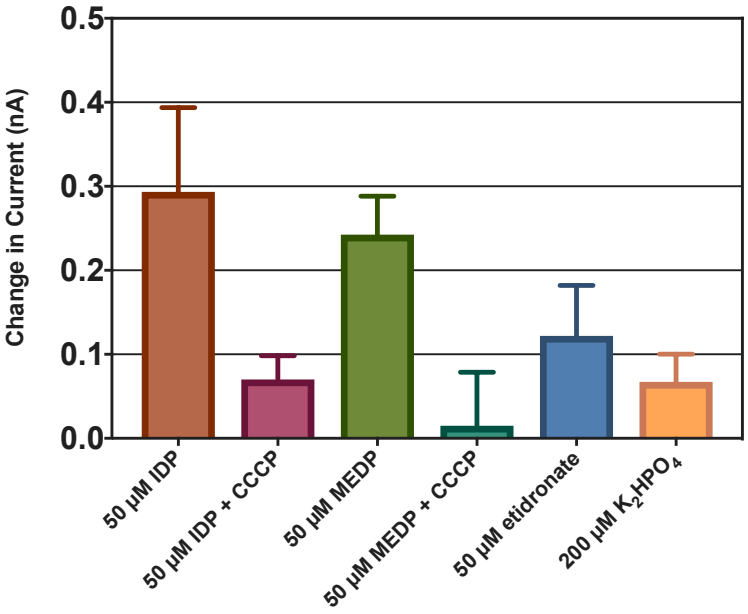


Figure 5.13: Effect of Inhibitors on the Proton Pumping Signal of VrPPase.
Effect of various inhibitors on the proton pumping signal generated in VrPPase following their respective addition to the activating buffer at the 1 second time point of SURFE2R experiments. The change in current is determined by subtracting the maximum peak value obtained from proton pumping measurements from the baseline prior to addition of the activating buffer. All results are from multiple replicates carried out on multiple sensor chips and error bars denote standard deviation.

5.4. Structural Comparison of M-PPases

5.4.1. Comparing all Crystal Structures of TmPPase and VrPPase

5.4.1.1. Investigating Hydrogen Bonding Patterns Between Catalytic States

Investigations were carried out between all the available crystal structures of M-PPases to study the differences and changes between different catalytic states. The TmPPase and VrPPase structures were analysed to investigate changes in the hydrogen bonding patterns and in the helical geometry. The webserver, HBplot (Bikadi *et al.*, 2007), identifies all the hydrogen bonding partners by analysing distances between atoms of a PDB structure.

At lower resolutions (> 3.0 Å), assignment of hydrogen bonding patterns may not be accurate, as atomic positions can vary quite significantly. However, the three main structures tested were the TmPPase:CaMg structure (2.6 Å), VrPPase:Mg_5IDP (2.35 Å) and TmPPase:Mg_5IDP (3.5 Å). Assignment of hydrogen bonding patterns could be trusted for the TmPPase:CaMg and the VrPPase:Mg_5IDP structure as these were at a higher resolution. Structural comparisons between TmPPase:Mg_5IDP and VrPPase:Mg_5IDP (Chapter 4) have also shown that helices and key residue side chains were in highly similar positions, meaning that the hydrogen bonding patterns of TmPPase:Mg_5IDP could also be trusted.

The Hbplot server was used on all the M-PPase structures and showed how the hydrogen bonding pattern changed between different states of the catalytic cycle, most notably in helix 6. In the TmPPase:CaMg (resting state) structure, the region between $N^{6.44}$-$V^{6.48}$ formed a small section of π-helix, differing from the standard α-helical geometry elsewhere on the helix. In both of the IDP-bound structures, the geometry of this region had changed, now forming a standard α-helical pattern. This represents quite a significant change, since $D^{6.43}$ has been previously linked to coordination of the water nucleophile responsible for hydrolysis of the substrate (Tsai *et al.*, 2014).

5.4.1.2. Investigating Changes in Helical Conformations Using BENDIX

A plugin for the Visual Molecular Dynamics (VMD) software (Dahl *et al.*, 2012), Bendix, was used to further study the geometry of helix 6. This software calculates the degree of curvature in each helix of a protein from the Cα trace. In the resting state structure of TmPPase, the region around $D^{6.43}$ was highly distorted with a bend angle of 22˚ (next

highest peak 14˚) (Figure 5.14), dropping to 11˚ in the IDP-bound structure and remaining at this level in both product-bound structures. In VrPPase, the available structures (IDP and product-bound) all show bend angles around 8-12˚, in line with the TmPPase structures. The absence of a resting state structure for VrPPase limits the comparisons that can be made just for the VrPPase structures, but the similarities between the TmPPase:Mg_5IDP and VrPPase:Mg_5IDP structures does not rule out the possibility that the region around $D^{6.43}$ undergoes a similar change in VrPPase as well. This decrease in helical curvature implies that there is a release of torsional strain energy in this region of helix 6 upon substrate binding, causing the helix to adopt a more energetically favourable conformation. To try to understand this change, additional analysis was carried out on the structures, paying close attention to the region of helix 6 around $D^{6.43}$.

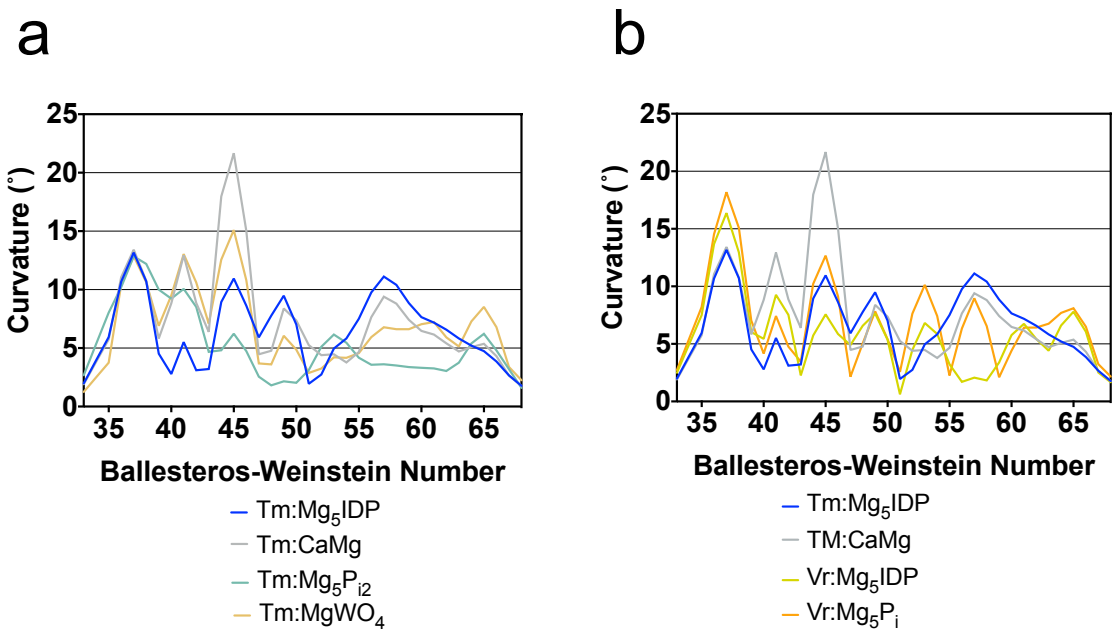


Figure 5.14: Analysis of Helix 6 Curvature Between Catalytic States of M-Ppases.
BENDIX results for structural analysis of helix 6 between (a) all four states of TmPPase from current crystal structures and (b) resting state and IDP-bound state of TmPPase alongside IDP-bound and product-bound states of VrPPase. All graphs depict the degree of helix curvature against the residue number in helix 6.

5.4.1.3. Structural Comparisons to Investigate Mechanisms of the Catalytic Cycle

In previous studies, differences had been noted between TmPPase resting state structure and the IDP-bound structure of VrPPase. In particular, a "downwards" motion of helix 12 upon substrate binding, leading to the breakage of a set of interactions at the base of the hydrolytic centre (Tsai *et al.*, 2014). However, since each structure was from a

different species and type of M-PPase (Na^+-PPase and H^+-PPase), it was not known how significant these differences were. The newly solved TmPPase:Mg_5IDP structure provides the means for an accurate comparison between the resting state and substrate-bound states from the same type of M-PPase. Comparison of both TmPPase structures shows that the 2 Å "downwards" motion of helix 12 still occurs upon substrate binding (Figure 5.15). This is likely due to the presence of several aspartate residues on helix 12 at the active site, which move to interact with the substrate or inhibitor upon binding, drawing the entire helix down by around 2 Å. Moreover, the TmPPase:Mg_5IDP structure confirms that the movement of helix 12 breaks the interactions between $K^{12.50}$ and $D^{6.43}$ at the base of the substrate-binding pocket in the hydrolytic centre.

In TmPPase:CaMg, $K^{12.50}$ interacts with $D^{6.43}$ and $D^{16.39}$ on helices 6 and 16 (Figure 5.15b), the same can be seen in VrPPase:Mg_3P_{i2} (Figure 5.15d). Upon substrate binding, the motion of helix 12 moved the lysine out of interaction range with $D^{6.43}$ (Figure 5.15). This led to a conformational change in $D^{6.43}$, which has flipped around towards the base of the active site and the bound substrate. This motion is also seen in VrPPase:Mg_5IDP, where the resolution was sufficient to resolve water molecules in the channel and around the bound substrate. In VrPPase:Mg_5IDP, $D^{6.43}$ and $D^{16.39}$ both interact with a water molecule that is in turn interacting with the bound substrate (Figure 5.15d). This represents an important aspect of the hydrolytic mechanism as this water molecule was previously proposed as the nucleophile that hydrolyses the substrate. Both $D^{6.43}$ and $D^{16.39}$ are required for activity, with $D^{6.43}$G mutations generating a completely inactive enzyme in *Streptomyces coelicolor* M-PPase (Hirono *et al.*, 2007). As such, the motion of helix 12, which allows $D^{6.43}$ to flip round and form interactions with the substrate, appears to be a vital mechanism to prepare the enzyme for hydrolysis to occur.

The altered conformation of $D^{6.43}$ upon substrate binding potentially explains the mechanistic basis behind substrate specificity in these proteins. As shown in the $SURFE^2R$ experiments, the structure of the substrate or inhibitor is vital to the closing of the 5-6 loop. Only the substrate or structurally similar inhibitors, such as IDP and MEDP, were able to bind and allow the protein to continue to pump protons, implying the 5-6 loop had closed. This could be explained by the resulting movement of helix 12, which would only occur if the substrate had bound correctly. Without this movement, the region of helix around $D^{6.43}$ would not rotate around from π-helical geometry to α-helical

geometry, $D^{6.43}$ would not coordinate the nucleophilic water and hydrolysis would therefore not be possible.

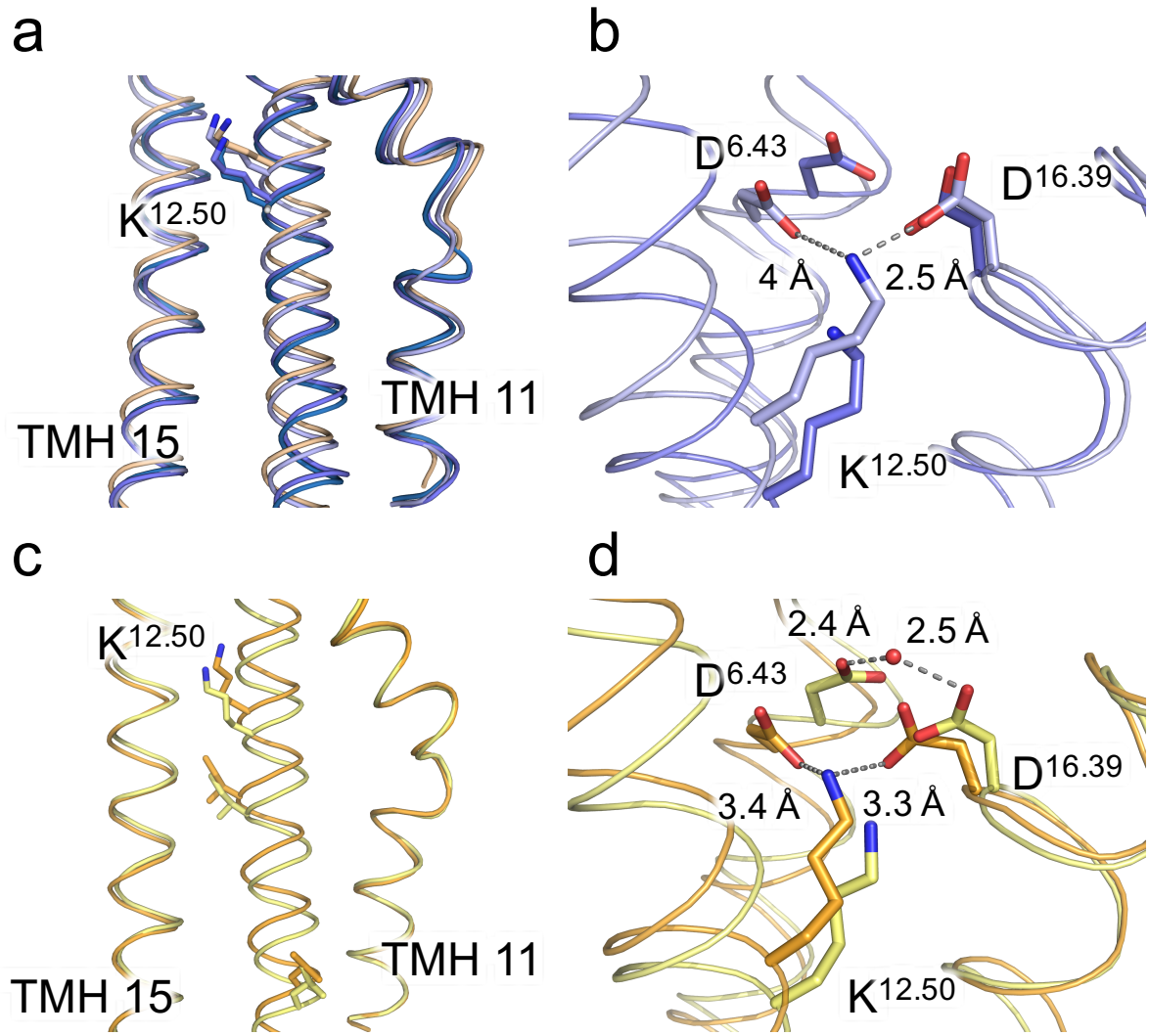


Figure 5.15: Structural Analysis of Hydrolytic Mechanisms in M-Ppases.
Structural comparisons between (a) helix 12 conformations of IDP-bound (blue), resting state (grey) and WO_4-bound (beige) structures of TmPPase (b) conformational states of $D^{6.43}$ in TmPPase:Mg_5IDP (blue) and TmPPase resting state (grey), (c) helix 12 conformations of VrPPase:Mg_5IDP (yellow) and VrPPase:$2P_i$ (orange) and (d) conformational states of $D^{6.43}$ in VrPPase:Mg_5IDP (yellow) and VrPPase:Mg_3P_{i2} (orange). Water molecules are displayed as red spheres.

The TmPPase:Mg_5IDP and VrPPase:Mg_5IDP structures, are globally very similar. They both represent the substrate-bound state of the catalytic cycle, in which the 5-6 loop is positioned above the hydrolytic centre sealing the cytoplasmic side of the channel. In both structures, there is a high degree of sequence conservation at the hydrolytic centre, with similar patterns of coordination seen for the Mg_5IDP complex (Figure 5.16). In VrPPase:Mg_5IDP, a potassium ion can be found in the hydrolytic centre along with water molecules that help to coordinate the Mg_5IDP complex. Given the positions of the

residues and Mg_5IDP complex in TmPPase:Mg_5IDP compared with those in VrPPase:Mg_5IDP, it is likely K^+ and the water molecules are present, but not observed at this resolution.

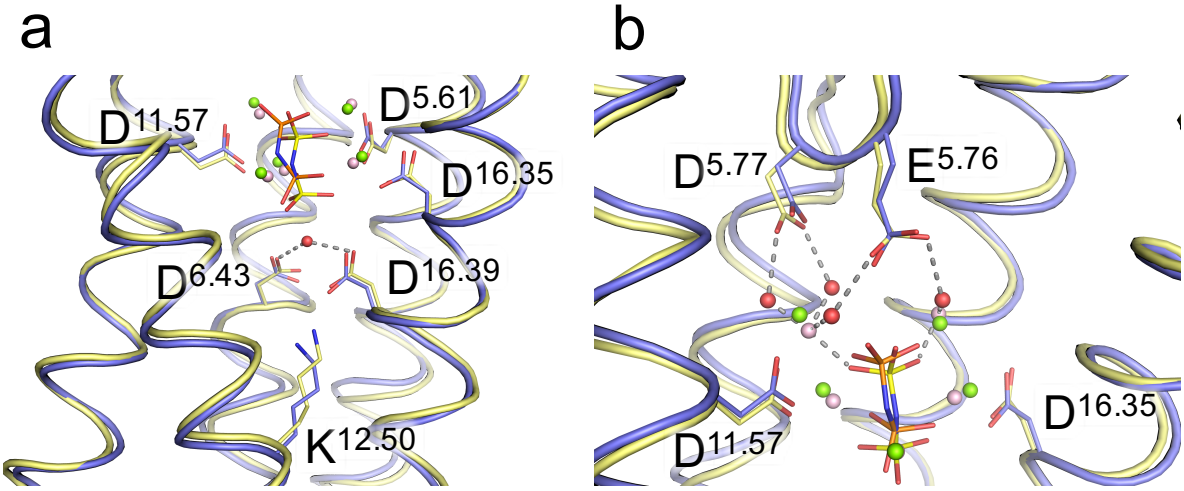


Figure 5.16: Structural Analysis of Hydrolytic Centre in M-PPases.
Structural comparisons of the hydrolytic centre showing (a) the interactions with the substrate, nucleophilic water and $D^{6.43}$ and (b) interactions between the substrate and residues of loop 5-6 between TmPPase:Mg_5IDP (blue carbon atoms, orange IDP and green Mg^{2+}) and VrPPase:Mg_5IDP (yellow carbon atoms, yellow IDP and pink Mg^{2+}). Water molecules are displayed as red spheres.

The similarity between the TmPPase and VrPPase structures was expected as the hydrolytic machinery was predicted to be the same in all M-PPases, despite differences in the type of M-PPase (H^+ or Na^+-pumping). VrPPase is a potassium-dependent H^+-PPase whereas the TmPPase is a K^+-dependent Na^+-PPase, with notable differences at the ion gate. A glutamate residue is located one helical turn down on VrPPase ($E^{6.57}$) compared with TmPPase ($E^{6.53}$) (Figure 5.17). The position of this glutamate was initially believed to play a role in the ion selectivity at the ion gate (Tsai *et al.*, 2014), being located at $E^{6.57}$ for all H^+-PPases and at $E^{6.53}$ for all Na^+-PPases. However, analysis of the various M-PPase sequences revealed that K^+-independent H^+-PPases (such as PaPPase) also had the glutamate located at position $E^{6.53}$, the same as the Na^+-PPases.

In the resting state TmPPase structure the ion gate is formed from interaction between $K^{16.50}$ and $D^{6.50}$, $E^{6.53}$ and $S^{6.54}$. However, this gate structure was different in the TmPPase:IDP structure (Figure 5.17) as $K^{16.50}$ has moved down and become less ordered, with no electron density visible for the NH_3^+ group. In its place, a sodium ion was bound around the ion gate (Figure 5.17), occupying the position of the NH_3^+ group of $K^{16.50}$.

$E^{6.53}$ appears to form part of the sodium binding site, coordinating Na^+ alongside $D^{6.50}$, $S^{6.54}$ and $D/N^{16.46}$ (Aspartate residue in TmPPase, but an asparagine in almost all other M-PPases). Previous mutagenesis experiments had linked $D^{6.50}$ to sodium binding and the structural information obtained from the TmPPase:IDP structure seems to back up this evidence. Since TmPPase is an Na^+-PPase, this sodium represents the ion bound at the ion gate prior to being pumped across the membrane. As a result of this information, residues surrounding $D^{6.57}$ in the VrPPase structure were checked to see if there was evidence of proton binding around this region.

In VrPPase:Mg_5IDP, $E^{6.57}$ has moved away from $K^{16.50}$ when compared with the product-bound structure (VrPPase:Mg_3P_i), in which $K^{16.50}$ directly interacts with $E^{6.57}$ (Figure 5.17b). This glutamate may have become protonated, leading to a loss of interactions with the NH_3^+ group of $K^{16.50}$ and instead forming new interactions with the backbone carbonyl of $K^{16.50}$ (Figure 5.17a). The difference between the two IDP-bound structures is due to cation specificity, with the charged sodium ion located above $K^{16.50}$ at the ion gate in TmPPase:Mg_5IDP and below $K^{16.50}$ in VrPPase:Mg_5IDP. These structures show two subtly distinct sub-states in which the ion or proton is passing through the ion gate prior to being pumped. These structures provide an interesting insight into the mechanism of ion translocation, particularly through the ion gate, and helps to provide more information on specific mechanisms occurring during the overall catalytic cycle that allowed for the model of the catalytic cycle to be updated (Chapter 6).

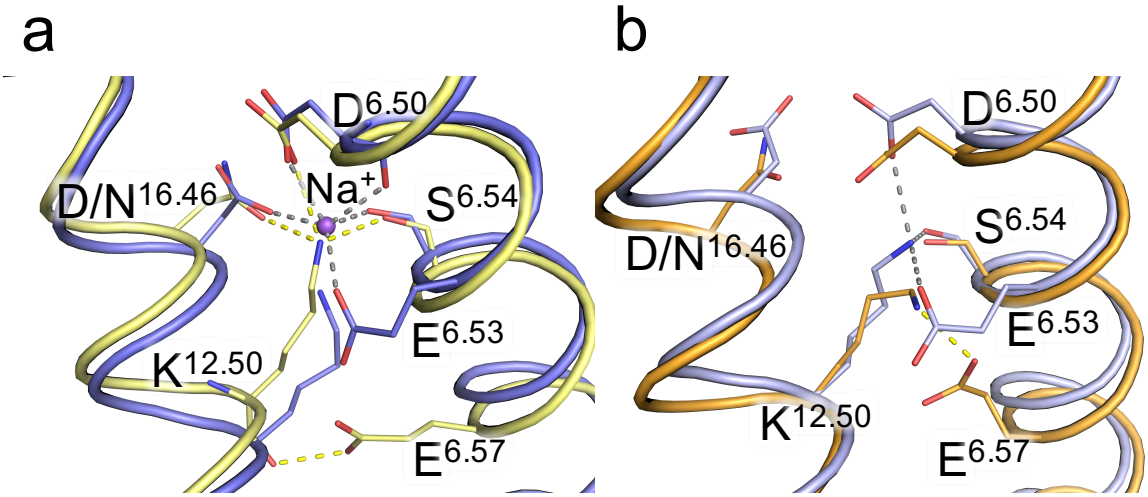


Figure 5.17: Structural Analysis of the Ion Gate of M-PPases.
Structural comparison of the ion gate region of (a) TmPPase:Mg_5IDP (blue) alongside VrPPase:Mg_5IDP (yellow) and (b) TmPPase resting state (grey) and VrPPase:Mg_4P_{I2} (orange).

5.5. Summary of Biochemical and Structural analyses of M-PPases

Through the work in this chapter, I was able to analyse all available structures of M-PPases, both those from the current study and those solved previously. By comparing the structures from different catalytic states, I was able to find specific information about the mechanisms of M-PPases, particularly relating to the hydrolytic and ion pumping machinery. I also collected biochemical data from tryptophan fluorescence and ion pumping experiments, providing further evidence to back up the structural data and help in the generation of the complete catalytic cycle model that will be discussed in chapter 6.

The ANS fluorescence data obtained showed the usefulness of this technique for studying conformational changes, but was not without its issues. The original data for the TmPPase protein showed a change in fluorescence signal with increasing inhibitor concentration. However, it was later determined that this change in fluorescence signal was generated by unfavourable interactions between magnesium and high concentrations of inhibitor, with no change in fluorescence signal detected at biologically relevant concentrations of IDP. When tested under optimal conditions, there was a clear change in fluorescence signal for both the VrPPase and PaPPase proteins with increasing concentrations of IDP. The VrPPase provided the greatest change in fluorescence, up to 10 %, which means it would be suitable for taking forward to stopped-flow experiments to measure conformational changes in real-time. This will be work carried out as part of the future directions of the project.

The SURFE^{2}R experiments were used to study ion pumping in VrPPase. The protein seemed to work well on the system, producing a repeatable signal up to 3 nA in the presence of substrate. Controls with no protein, without substrate and in the presence of ionophores confirmed that the signal was due to proton pumping and not an artefact of the technique itself. It was possible to detect a signal above background levels by substituting the substrate for various analogues, IDP and MEDP, which are both structurally similar to the pyrophosphate substrate. Inversely, etidronate, which contains an extra methyl and hydroxyl group, did not produce a signal above background. Ionophores were used to confirm that this signal was caused by proton pumping and

provided preliminary data to suggest that ion pumping occurs before hydrolysis during the catalytic cycle.

Analysis of the current M-PPase crystal structures also highlighted several interesting features that provided insight into the mechanisms of these proteins. The main mechanism identified involves the priming of the enzyme for hydrolysis, which occurred upon substrate binding via the downwards motion of helix 12 by around 2 Å. This removed interactions between $K^{12.50}$ and $D^{6.43}$ at the base of the active site, released torsional strain energy around a single turn of helix 6 and changed the hydrogen bonding pattern around $D^{6.43}$ from π-helical geometry to α-helical geometry. In doing so, the position of $D^{6.43}$ changed to coordinate the nucleophilic water alongside $D^{16.43}$. In this way, the binding of substrate can prime the enzyme for hydrolysis. This mechanism also accounts for substrate specificity and provides an explanation of how this might be achieved in M-PPases.

The structural analysis also identified differences between the two substrate-bound states from VrPPase and TmPPase. There was evidence to suggest that a sodium ion was bound at the ion gate in TmPPase:Mg_5IDP, where it had disrupted the ion gate architecture by displacing $K^{16.50}$. It was proposed that $E^{6.57}$ in VrPPase:Mg_5IDP had become protonated, meaning the ion being pumped is located above $K^{16.50}$ in TmPPase:Mg_5IDP, but below $K^{16.50}$ in VrPPase:Mg_5IDP. These structures represent two distinct sub-states in which the ion is transitioning through the ion gate prior to being pumped. Using all of this information, it is possible to generate a model of the complete catalytic cycle, which will be discussed in greater detail in chapter 6.

Chapter 6:
Overall Conclusions, Discussion and Future Directions

6.1. **Discussion and Key Findings from the Project**

Overall, I have been able to achieve the aims of my project, to update the model of the complete catalytic cycle of M-PPases using new structural and biochemical information. As shown in chapter 3, both TmPPase and PaPPase were expressed in yeast following modified versions of the original "hot-solve" solubilisation and purification method (López-Marqués *et al.*, 2005). TmPPase was expressed using the modified protocol from previous crystallisation experiments (Kellosalo *et al.*, 2011), but PaPPase required optimisation to obtain sufficient yields of pure protein for crystallisation experiments. Several constructs were tested, based on the presence of linkers between the His-tag and protein, to test the binding and elution of the protein from the Ni-NTA column, and signal sequences from *Trypanosoma cruzi*, to try and improve the yield. Ultimately, the presence of a "GG" linker between the His-tag and protein improved the yield and so was used for all subsequent experiments. The use of selective SCD-Leu media showed the highest level of expression, as non-selective media led to the removal or inactivation of the plasmid from the yeast during expression. Solubilisation and purification conditions were also optimised, based on the differing growth and activity temperatures of PaPPase (81 ˚C) compared with TmPPase (71 ˚C).

Crystallisation experiments were carried out using PaPPase purified in DM, as this had yielded the best crystal in previous experiments (Kellosalo *et al.*, 2013). Initial crystal hits diffracted to around 6.5 Å, providing only limited information about the arrangement of helices in the dimer. Screening different detergents at different concentrations, during purification of PaPPase did not improve the diffraction of these crystals. Initial dehydration experiments also did not improve the diffraction of the crystals. However, further dehydration experiments, in which the crystals were dehydrated in the loops during harvesting and freezing, yielded diffraction data to around 4.3 Å. Using this technique, it was possible to reliably obtain crystals diffracting to between 4.2 and 4.5 Å. These data provided useful information about the helical conformations and placement of some bulkier side chains was possible (Figure 4.9). However, key information about the mechanism of potassium independence, $K^{12.46}$, was not resolvable as there was no electron density for the side chain of this residue.

During early rounds of data processing in XDS, I noticed that the first few hundred images from each dataset had data going to a higher resolution (~ 3.8 Å). However, obtaining this higher resolution data led to poor completeness and redundancies for each individual dataset. Since each dataset was in the same space group with very similar unit cell parameters, these seven datasets were processed using only the first 100-400 diffraction images (10-40˚ rotation) and were then merged to obtain a single data file. The resulting dataset was at 3.8 Å resolution and the electron density maps provided much more information. The inhibitor IDP could be seen in the active site, with density to support many of the complexed magnesium ions (Figure 4.12). There was also density to fit more side chains compared with the 4.3 Å datasets (Figure 4.11) and there was also limited density to suggest a conformation for $K^{12.46}$, which has been linked to the K^+-independence mechanism of M-PPases (Figure 4.16).

The new PaPPase:Mg_5IDP structure also showed many similarities to the IDP-bound structures from VrPPase and TmPPase, solved previously and in the current study, respectively. There were several differences in helical positions in PaPPase:Mg_5IDP, including helices 13 and 14, which suggests a possible mechanism for K^+-independence alongside the role of $K^{12.46}$ (Figure 4.14). Moreover, mutational studies in PaPPase revealed that mutation of both $K^{12.46}$ and $T^{12.49}$ to alanine could generate a K^+-dependent variant (Figure 3.15). These findings differ from previous mutagenesis experiments in which a K^+-dependent M-PPase from *C. hydrogenoformans* could be mutated into a potassium-independent variant solely via $A^{12.46}K$ mutation, with $A^{12.49}T$ only having an effect on the affinity for K^+ (Belogurov & Lahti, 2002). PaPPase remained K^+-independent with only the $K^{12.46}A$ mutation, implying that either Na^+ can substitute for K^+ in H^+-PPases and not just in Na^+-PPases or that the mechanism of K^+-dependence is more complicated than previously thought, potentially involving more residues than $K^{12.46}$. The resolution of the PaPPase:Mg_5IDP structure provides some useful structural information to start to identify these other factors, but higher resolution information and structures of these mutants will be needed to fully confirm these findings.

The new TmPPase:Mg_5IDP and TmPPase:$MgWO_4$ structures represented unique catalytic states of TmPPase and provide complementary information to the previously solved structures of TmPPase and VrPPase. Comparisons between the TmPPase:Mg_5IDP and VrPPase:Mg_5IDP structures reveals a high degree of similarity

between the helices and side chain conformations of many of the key residues, with the closure of the 5-6 loop visible in both structures. The composition of the hydrolytic centre, including bound Mg_5IDP, is highly similar and provides evidence to suggest that the hydrolytic mechanism is identical in both proteins, despite differences in species and cation specificity (Na^+ and H^+) (Figure 4.3). The only notable differences exist at the ion gate, in which a sodium ion is bound above $K^{16.50}$ in TmPPase:Mg_5IDP, with the binding site located around the previously identified $D^{6.50}$ residue and $E^{6.53}$ (Figures 4.5). This glutamate is located one helical turn down in VrPPase ($E^{6.57}$) and in the IDP-bound structure may be protonated, meaning the cation is located below $K^{16.50}$. This represents a subtle shift in the cation position across the ion gate and shows that these two structures represent two distinct sub-states of the substrate-bound state of M-PPases (Figure 4.6). The TmPPase:$MgWO_4$ structure represents an analogue of the single-product bound state and is the last major state of the catalytic cycle before the return to the resting state. The conformations of the helices are more similar to the resting state structure than to the substrate-bound or two-product bound states, with the 5-6 loop being highly disordered.

To complement the structural data, I was also able to collect biochemical data in order to study conformational changes in various M-PPases. I was trying to identify different catalytic states, backing up the structural data on existing states and identifying previously unknown states, such as the high-energy transition state. My studies on the tryptophan fluorescence of PaPPase (Figure 5.6) and VrPPase (Figure 5.7) revealed that conformational changes could be detected following the titration of IDP. The positions of tryptophan residues in TmPPase meant that changes in conformation could not be detected using this method as they were in regions of the proteins that do not undergo significant conformational changes between the TmPPase:CaMg and TmPPase:Mg_5IDP structures (Figure 5.5). These represent the largest known conformational change during the catalytic cycle.

The change in fluorescence signal detected for PaPPase was around 5 %, which is too small to take forward to stopped-flow experiments as this signal would likely be lost to background noise. However, the signal obtained for VrPPase was around 8-10 %, which makes it suitable for future stopped-flow fluorescence experiments. Due to time constraints, these experiments could not be carried out as part of this project. However,

these would involve taking real-time measurements in the millisecond range using titrations of the PP_i substrate. The hope would be to try and map conformational changes to different states of the catalytic cycle and potentially identify new states that could not be identified through standard structural biology techniques. This will be discussed in more detail in section 6.2.2.

To investigate ion pumping in M-PPases and try to understand how this fits into the model of the catalytic cycle, experiments were carried out on VrPPase using the Nanion SURFE^{2}R N1. Initial experiments using the PP_i substrate revealed a proton-specific signal up to 3 nA when 50 μM PP_i was added to the activating buffer during the experiments (Figure 5.9). When PP_i was replaced with a variety of non-hydrolysable inhibitors in the activating buffer, a proton-specific signal was still observed. The inhibitors, IDP and MEDP, produced a signal of 0.4 nA, significantly above the background signal when phosphate was used (Figures 5.11 & 5.13). Both of these inhibitors are close structural analogues of PP_i (Table 5.1) and would therefore allow loop 5-6 to close over the active site. The use of the ionophores, gramicidin and CCCP, reduced this signal to the background level of 0.1 nA, implying the signal was generated via proton pumping and not any other factor (Figure 5.13). Inversely, the inhibitor, etidronate, when added to the activating buffer in place of PP_i, did not produce a signal (Figure 5.11). This inhibitor has structural differences from PP_i, with the addition of a hydroxyl and methyl group that would form a steric clash with several key residues and magnesium ions in the active site, preventing correct closure of the 5-6 loop (Figure 5.14). Overall, these experiments provided preliminary evidence that ion pumping precedes hydrolysis during the catalytic cycle as proton pumping was observed in the presence of non-hydrolysable inhibitors.

6.1.1 Updated Model of the Complete Catalytic Cycle

Using all the available data, both from the previous crystal structures and from the current studies (structural and biochemical), it has been possible to create a complete model of the catalytic cycle. Several attempts had previously been made to do so, using the limited structural data available at the time. Three main mechanisms had been proposed for the catalytic cycle (Figure 1.11). The first was the "binding-change" mechanism, in which the binding of substrate drives a series of conformational changes that lead to ion pumping, the resulting charge difference between the ion gate and hydrolytic centre then drives hydrolysis followed by sequential product release (Kellosalo *et al.*, 2012). In the second mechanism, the substrate is hydrolysed upon binding, with the resulting conformational changes from this leading to ion pumping via a Grotthuss chain mechanism (Lin *et al.*, 2012). The key difference between these mechanisms is the order of ion pumping and hydrolysis upon substrate binding. The third mechanism has substrate hydrolysis occurring first, with the resulting proton being directly pumped (Baykov *et al.*, 2013). This mechanism is unlikely to be correct as it does not account for sodium pumping, and the mechanism of each is predicted to be very similar (Tsai *et al.*, 2014). The evidence provided in this study backs up the "binding change" mechanism (Figure 1.11b), in which ion pumping precedes hydrolysis in the catalytic cycle. Using all the available information, an updated model of the complete catalytic cycle was generated (Figure 6.1).

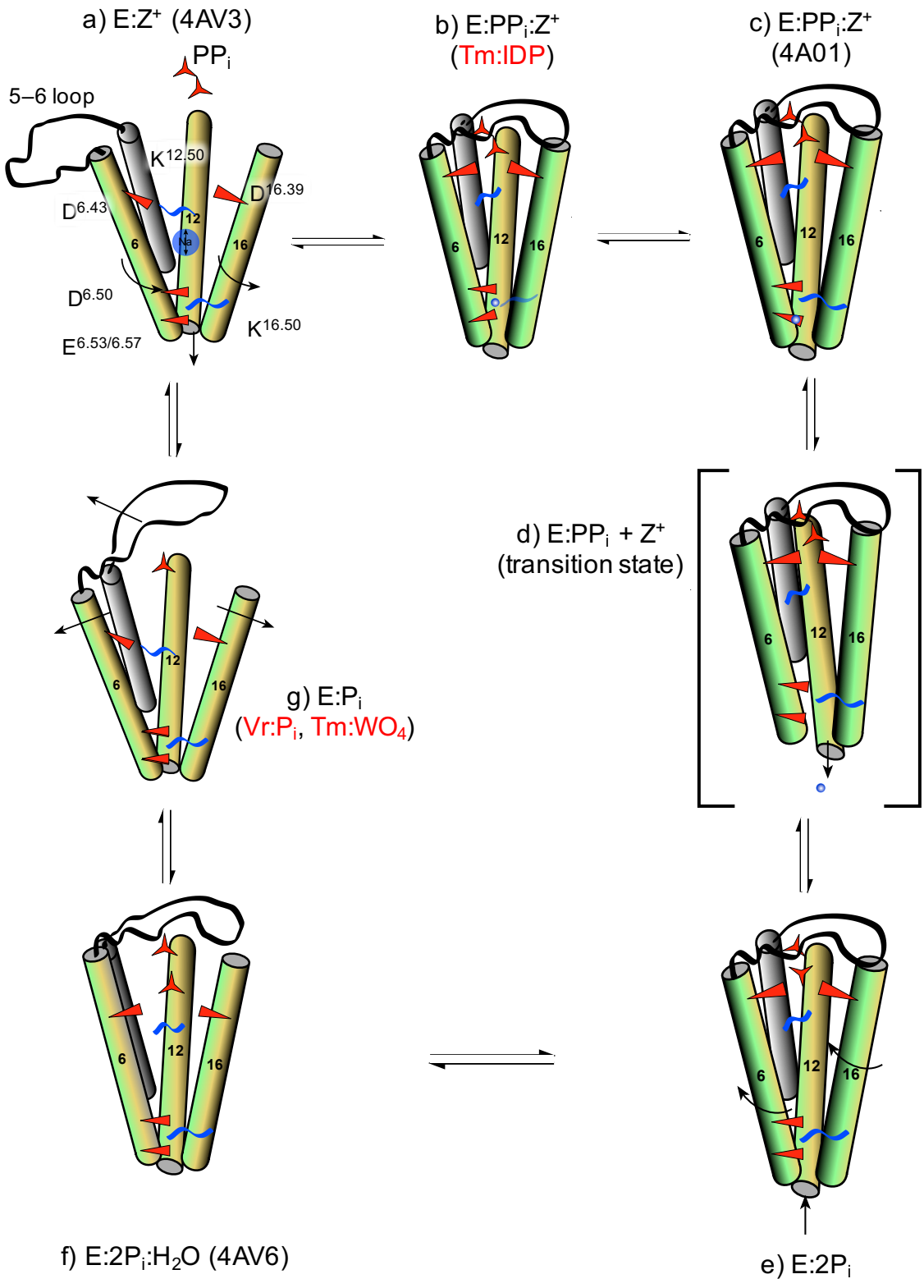


Figure 6.1: Overview of the Complete Catalytic Cycle Model for Membrane Pyrophosphatases. *Simplified view of a generalised membrane pyrophosphatase during the catalytic cycle. For clarity, only key helices and residues are shown and labelled. PDB IDs are given for previously solved structures and structures named in red represent structures solved as part of this project and related publication (K. M. Li et al., 2016).*

The first state of this cycle is the resting state (Figure 6.1a, Figure 6.2), obtained from the TmPPase:CaMg structure (PDB ID: 4AV3). In this state, the 5-6 loop is highly disordered and is not interacting with the hydrolytic centre, leaving the cytoplasmic side of the channel open. The ion gate is closed and no cation can be observed in this region. $D^{6.43}$ interacts with $K^{12.50}$ at the base of the hydrolytic centre and helix 6 is under torsional strain at this position. Substrate binding (Figure 6.1b) involves the new TmPPase:Mg_5IDP structure. Substrate binds at the hydrolytic centre leading to the closure of loop 5-6 via the interactions of $E^{5.76}$ and $D^{5.77}$ with the substrate complex (Figure 5.17). Additionally, several aspartate residues on helix 12 start to form interactions with the substrate complex, leading to a downwards motion of this helix by around 2 Å away from the cytoplasmic side. This movement breaks the interactions between $D^{6.43}$ and $K^{12.50}$, releasing the torsional energy in this region of helix 6 (Figure 6.2). This changes the conformation of the helical region around $D^{6.43}$ from π-helical geometry to α-helical geometry. This results in both $D^{6.43}$ and $D^{16.39}$ coordinating the nucleophilic water (Figure 5.16, Figure 6.2). In doing so, the enzyme is now primed for hydrolysis, but another factor is needed for this to occur. The cation (Na^+ or H^+) has passed along the charged coupling funnel and is now bound at the ion gate, above $K^{16.50}$

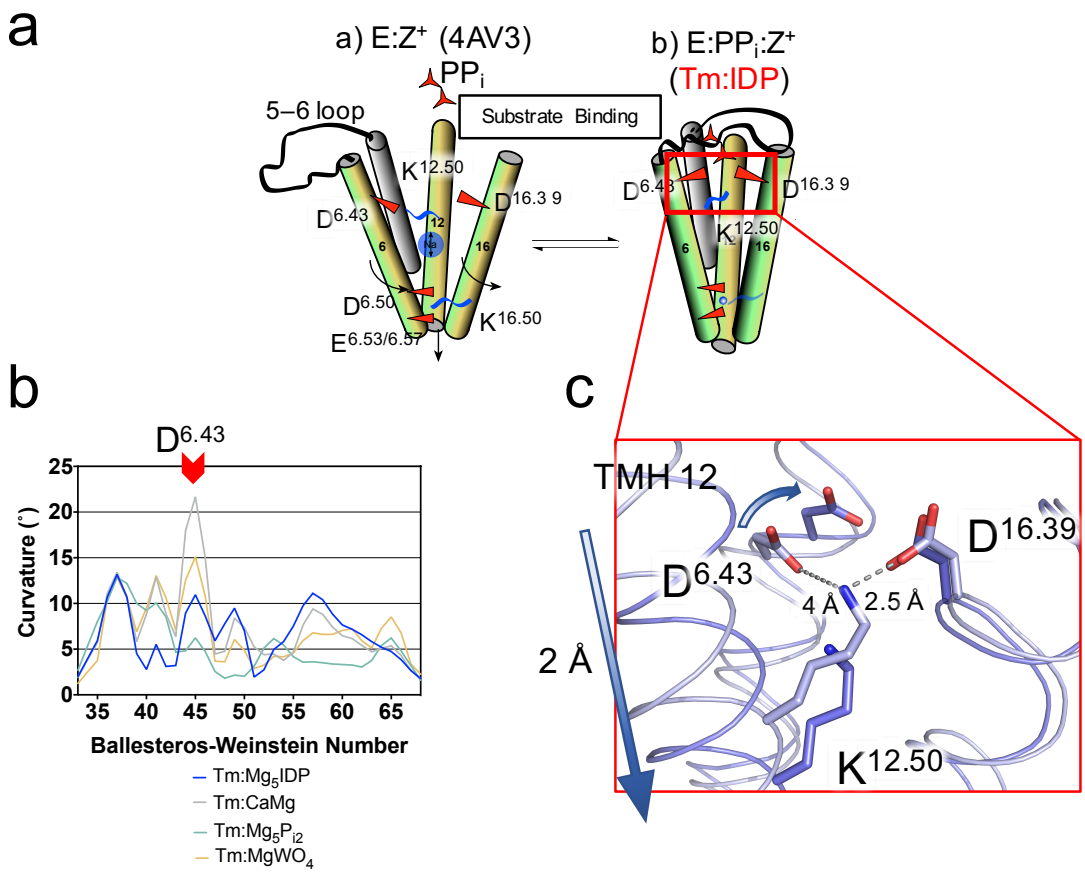


Figure 6.2: Overview of Substrate Binding Step of the M-PPase Catalytic Cycle.
Substrate binding step of the M-PPase catalytic cycle showing the triggering of the "molecular mousetrap" following substrate binding. (a) excerpt from Figure 6.1 showing only the first two states representing substrate binding. (b) Bendix plot for helix 6 of TmPPase showing the degree of helix curvature for the four main catalytic states, including the resting state (grey) and substrate-bound state (blue). (c) Close-up structural view of the base of the hydrolytic centre comparing the interactions of $K^{12.50}$ between the resting state (grey) and substrate-bound state (blue), showing the "molecular mousetrap" mechanism formed by $D^{6.43}$.

The next step in the catalytic cycle involves a subtle change in the position of the cation (Figure 6.1c, Figure 6.3). The next available structure is the VrPPase:Mg_5IDP (PDB ID: 4A01), in which the cation at the ion gate has moved from above to below $K^{16.50}$. This represents a series of sub-states in which the cation transitions through the ion gate prior to being pumped. The residues in the hydrolytic centre, including the positioning of the substrate complex, $D^{6.43}$ and loop 5-6 are all in very similar positions between the two IDP-bound structures (Figure 4.3, Figure 6.3).

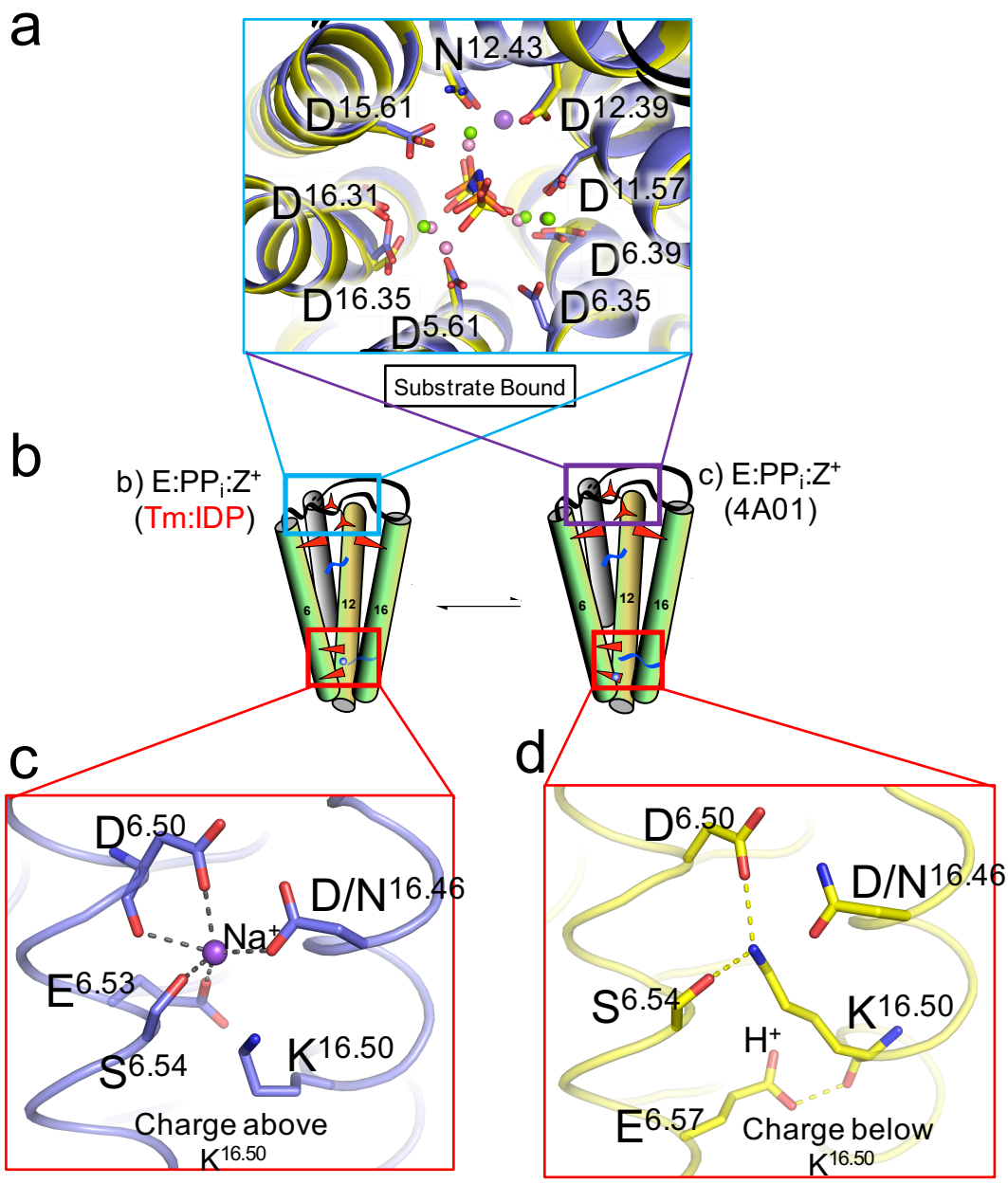


Figure 6.3: Overview of the Substrate-bound Sub-states of the M-PPase Catalytic Cycle.
Comparison of the structural differences between the two substrate-bound structures of M-PPase. (a) comparison of the hydrolytic centre of the substrate-bound structures of TmPPase:Mg_5IDP (blue carbon atoms, orange IDP and green Mg^{2+}) and VrPPase:Mg_5IDP (yellow carbon atoms, yellow IDP, pink Mg^{2+}, purple K^+). (b) excerpt from Figure 6.1 showing only the two substrate-bound states of M-PPases. (c) Overview of the TmPPase:Mg_5IDP ion gate region, showing Na^+ as a purple sphere, positioned above $K^{16.50}$. (d) Overview of the VrPPase:Mg_5IDP ion gate showing the possible protonated glutamate ($E^{6.57}$), positioned below $K^{16.50}$.

The next stage of the catalytic cycle is the high-energy transition state. It is known that both ion pumping and hydrolysis occur during this state, but all available structural data could not predict in which order these events occurred (Figure 6.1d). No structural information was available for this state as it represented the gate-open state, which occurs at time-scales too quick to capture using standard crystallographic techniques. The Nanion SURFE^{2}R data revealed that ion pumping occurs before hydrolysis during the catalytic cycle, providing evidence to back up the "binding change" mechanism (Figure 1.11b, Figure 6.4). A possible mechanism for this is via additional downwards motion of helix 12, leading to opening of the ion gate, cation release and immediate closure of the ion gate as helix 12 moves back up to prevent backflow of ions. This creates a charge difference between the ion gate and the hydrolytic centre, leading to a movement of cations down the coupling funnel to replace the lost charge at the ion gate. This leads to the activation of $D^{6.43}$ and $D^{16.39}$ at the base of the hydrolytic centre, in turn activating the water nucleophile and leading to hydrolytic attack on the substrate.

Following substrate hydrolysis is the immediate product bound state (Figure 6.1e), in which the positions of the two phosphate molecules very closely matches those of the bound substrate, with the electrophilic phosphate positioned down towards $D^{6.43}$ and $D^{16.39}$. The overall structure is still highly similar to the substrate-bound state, with the only exception being the presence of the product in place of the substrate. The next stages of the catalytic cycle (Figure 6.1e-g) involve opening of the 5-6 loop and sequential product release. The first stage of this involves the dilation of the helices around the cytoplasmic side of the channel, this allows the two phosphate molecules to alter position within the hydrolytic centre. Loop 5-6 then opens and the two phosphate molecules are released sequentially, with the electrophilic phosphate being released first, followed by the leaving group phosphate. This returns the protein to the resting state and allows the cycle to continue.

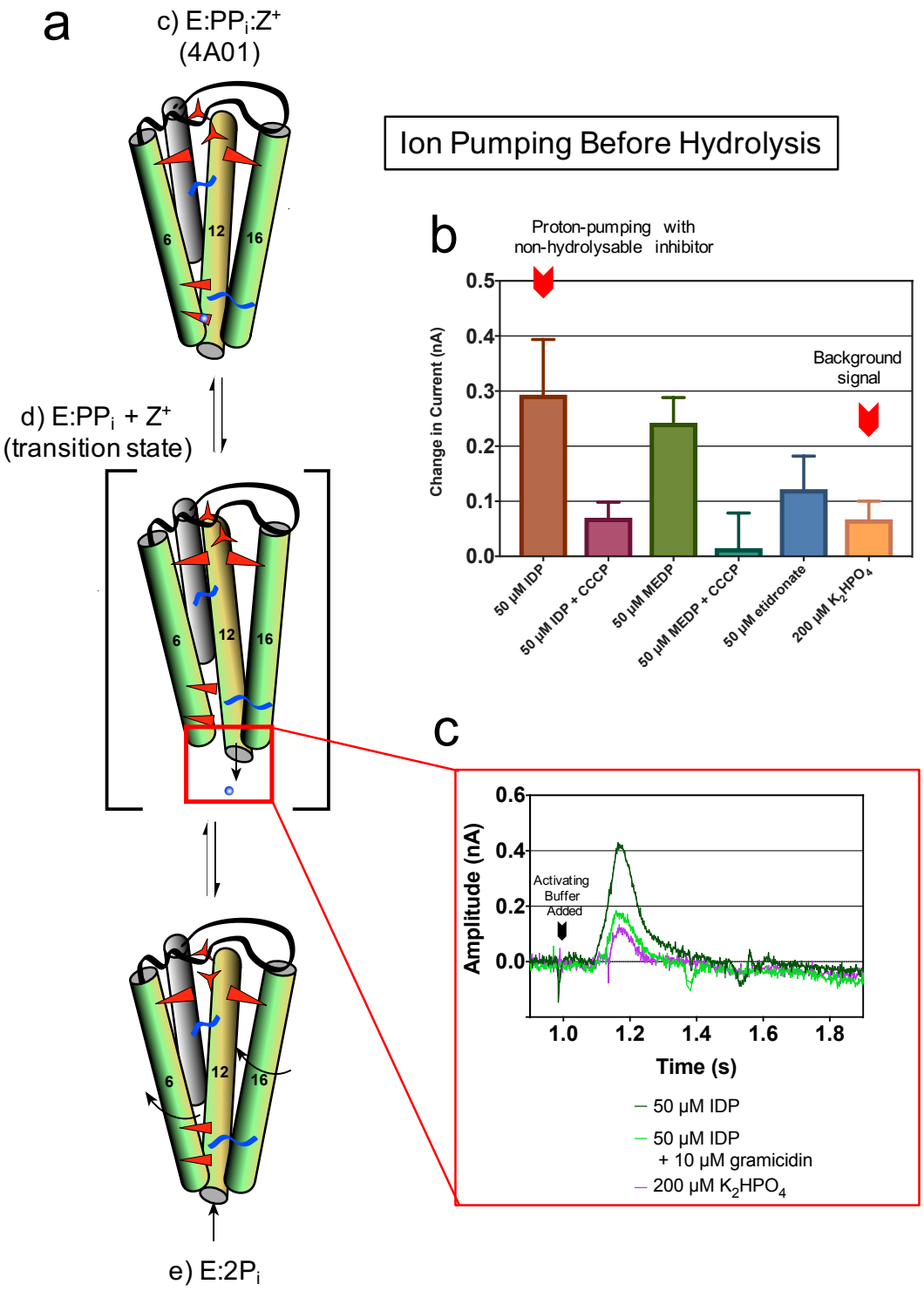


Figure 6.4: Overview of the Transition State of the M-PPase Catalytic Cycle.
(a) Excerpt from Figure 6.1 showing only the three states involved in the hydrolysis and ion pumping stages of the M-PPase catalytic cycle, the VrPPase:Mg_5IDP structure, the proposed high-energy transition state and the immediate-product bound structure. (b) SURFE^{2}R data showing the change in current (denoting a proton-pumping signal) when the listed inhibitors were added to the activating buffer for SURFE^{2}R experiments of VrPPase proteoliposomes. Results are from replications of multiple samples carried out on multiple sensors, error bars denote standard deviation. Red arrows indicate the proton-pumping signal observed following addition of the non-hydrolysable inhibitor, IDP, and the phosphate control to the activating buffer. (c) Proton-pumping signal generated when IDP was added to the activating buffer during the SURFE^{2}R experiments on VrPPase.

6.1.2 Summary of the Catalytic Cycle Model

The updated model of the catalytic cycle provides the first evidence to suggest a specific order of events, notably the occurrence of ion pumping before hydrolysis, backing up the previously proposed binding change mechanism (Kellosalo *et al.*, 2012). However, this information leaves several very important questions unanswered. The SURFE^{2}R data suggests that a single-turnover event can occur, in which a single ion can be pumped per protomer in the presence of a non-hydrolysable inhibitor, such as IDP. This would require a complex series of energy transfer and storage steps, as ion pumping can still occur without hydrolysis. It is possible that the energy required for ion pumping to occur is stored in a series of conformational changes within the protein, such as the torsional energy that is released during the triggering of the "molecular mousetrap" of D$^{6.43}$ upon substrate binding. Additionally, the larger movements of helix 12, which moves down by around 2 Å during substrate binding, may provide an alternative mechanism of transferring energy required for ion pumping to occur before hydrolysis.

Once ion pumping has occurred, it is the difference in charge between the ion gate and the hydrolytic centre that draws cations down the coupling funnel. The hydrolytic machinery has already been primed through the triggering of the "mousetrap" model and therefore the charge difference alone can lead to hydrolysis of the substrate. The release of energy from hydrolysis is then used to reset the "mousetrap" mechanism, altering the conformation of helix 12 and the region around D$^{6.43}$ to prime the "mousetrap" for binding of the next substrate. The energy is therefore stored in these energetically unfavourable conformational changes and in the process, the phosphate products are released when loop 5-6 opens.

The model does not account for any potential cross-dimer effects that may influence the cycle. It is currently not known if allosteric regulation of the cycle, in which one protomer is activated by events from the other protomer. This could have significance for the dual-pumping M-PPases, where the mechanisms of pumping protons and sodium ions has not been determined and allosteric regulation may play a part in determining which cation is pumped. This would be difficult to investigate as it is currently not possible to measure the pumping of protons or sodium ions from individual protomers. The crystal structures show that each protomer is in an identical state, but these occur

under saturating conditions of substrate or product. The location of the C-terminus near to the dimer interface provides suggests that these residues might work via stabilising the formation of the dimer. This could explain why the C-terminal truncation mutant is inactive, as the dimer may not form correctly. If so, this might suggest that there is some degree of allosteric effect from the dimer, which is unable to occur in the C-terminal truncation mutant.

The new structures and the updated model do not provide any insight into the mechanism of ion selectivity. Previously this was believed to be related to the position of a "semi-conserved" glutamate ($E^{6.53/6.57}$) (Tsai *et al.*, 2014) that is different in K^+-dependent H^+-PPases and Na^+-PPases. However, sequence alignments show that this glutamate is in the same position for Na^+-PPases as it is for K^+-independent H^+-PPases, ruling this residue out as a key selector for the cation. In the TmPPase:Mg5IDP structure, $E^{6.53}$ forms part of the sodium ion binding site, coordinating the bound sodium ion. The low resolution of the PaPPase structure (3.8 Å) means that the side chain conformation of $E^{6.53}$ cannot be determined and so nothing is known about its role in K^+-independent H^+-PPases. Scans of the sequence alignment between the various types of M-PPase have also not provided any insight into the ion selectivity of M-PPases.

6.2 Future Directions for the Project

The work carried out as part of this thesis has provided updated structural and functional data that has allowed me to generate an updated model of the catalytic cycle. This covers previously uncharacterised states of the cycle, such as the high-energy transition state and provides initial structural information to start analysing K^+-stimulated specificity in M-PPases. However, there is still much work to be done on M-PPases to fully understand these proteins. The model of the catalytic cycle could be refined using higher resolution structural data, as well as biochemical and computational data from a range of different types of M-PPase (H^+, Na^+ and Na^+/H^+-PPases). They will also allow us to better understand the specific mechanisms behind cation specificity and can also be used for drug design against protozoan parasites and some pathogenic bacteria.

6.2.1 Structural Methods to Improve our Understanding of M-PPases

Additional structural information from different types of M-PPase would help to refine the model of the catalytic cycle and allow for more detailed comparisons between Na^+-PPases and H^+-PPases or K^+-dependent and K^+-independent variants. These experiments involve trying to express, purify and crystallise the M-PPases from other species or involving other mechanisms of action. These include the dual-pumping M-PPases from certain prokaryotic species, which are capable of pumping both sodium ions and protons under certain conditions. A structural understanding of the mechanisms involved in pumping both cations would be invaluable to the M-PPase field and, potentially, to other dual-pumping transporters. The structures of protozoan parasite M-PPases or those from pathogenic prokaryotes would also provide extremely important information on these proteins and potentially provide information to assist in future drug design programmes.

Structural information can be used to help drive drug discovery. The layout of the hydrolytic centre, for example, can be used for docking studies of existing compounds or can be used to direct large-scale compound screens. Any that show activity can be crystallised and this information can then be fed back to medicinal chemists to alter and improve the drugs based on the available information. Work has already started on this using TmPPase as a model organism, given the high yield of expression, reliable purification method and high degree of sequence conservation between this and the M-PPase from pathogenic species. If the protozoan parasite M-PPases can be successfully expressed and purified, it may be possible to carry out inhibition assays on these proteins instead, to provide more reliable information on inhibition than testing on the TmPPase.

Higher resolution structural information on the PaPPase will also be useful for studying the K^+-dependence mechanism of M-PPases. There are several ways this might be achieved, including the additional screening of crystals in the current best crystallisation conditions. There is a great degree of variability in diffraction between crystals, even from the same drop, meaning that large amounts of crystals must be screened to find the best diffracting ones. Moreover, improvement in the crystal picking technique may increase diffraction of the crystals.

The best diffracting crystals were of a thin plate morphology, meaning that physical stress on the crystals during picking was enough to crack or destroy the crystals, leading to poorer quality diffraction and thus a lower resolution of the datasets. Possible solutions for this problem involve improving the technique and conditions when picking the crystals, such that the mechanical stress on the crystals is reduced and thus the final crystals are intact during data collection and so diffract better. This could be achieved using different types of loops, specific to plate crystals, which reduce the mechanical stress when picking the crystals. Alternatively, it might be possible to test room temperature diffraction using *in situ* data collection techniques available at synchrotrons. Another method would be to try and find alternative crystallisation conditions that produce crystals of a different morphology that diffract well and can survive the picking procedure. These alternative conditions could be different buffers, salts, precipitants, *etc* or be changes to the protein itself to better aid in crystal packing. Although the C-terminal truncation mutant in this current study yielded protein with almost no activity, it is possible that modifications that remove fewer residues from the C-terminal could have the desired effect of improving the stability of the protein, leading to better crystal contacts, without reducing the activity and structure of the protein. Alternative crystallisation methods could also be tested, such as using Lipidic Cubic Phase (LCP) to see if this could produce crystals that diffract better than those obtained in the current study.

There are also other structural biology techniques that might be useful to study the M-PPases. Electron microscopy can be used to study the structure of membrane proteins in a more native environment and without issues such as crystal contacts distorting the structure. In recent years electron microscopy has been able to reach higher resolutions around 4 Å, allowing for the study of specific mechanisms within proteins of smaller molecular weights, such as the TAP transporter at 166 kDa (Oldham *et al.*, 2016), which was solved via Cryo-EM. The issue with M-PPases is that they do not have any significant cytoplasmic domains and therefore are hard to differentiate from empty detergent micelles in EM. Combining Cryo-EM with the reconstitution of the protein in nanodiscs or amphipols means that EM could be used to study various M-PPases in a very native environment, potentially identifying features of the proteins that would not be visible in crystallography. Another method is to use LMNG during the purification, with no additional detergent added during the wash or elution steps on the Ni-NTA

column. This should produce detergent micelles containing the protein, but should limit or completely remove the empty detergent micelles from the sample, allowing for easier picking of protein-containing particles during EM. Since both TmPPase and PaPPase can be solubilised and purified in LMNG, this is a viable method to help overcome the limitations of differentiating particles containing empty micelles from those containing protein.

6.2.2 Biochemical Studies to Improve our Understanding of M-PPases

Alongside the structural work, there are additional biochemical techniques that could be tried to help better understand the M-PPases. The tryptophan fluorescence experiments provided some initial data to suggest that it would be possible to study conformational changes in the VrPPase protein, showing that titrations of the IDP produced a change in fluorescence signal around 8 %. If this sample was tested using a stopped-flow fluorimeter, it might be possible to observe conformational changes occurring over a millisecond timescale. These conformations could then be mapped to existing points on the catalytic cycle model or even identify states not seen from the crystal structures. This could also be a means of exploring the pre-steady state kinetics of the catalytic cycle, adding kinetic data to the model and potentially leading to more interesting observations about the opening of the ion gate.

Alongside the tryptophan fluorescence work, it would also be possible to test the M-PPases using small molecule Fluorescence resonance energy transfer (smFRET). Work is currently underway to study mutants of TmPPase using this system. This works by attaching fluorescence dyes to cysteine residues on the surface of proteins and measuring the distance between the dyes. If measured in the presence of substrate, it might be possible to observe conformational changes and, much like with tryptophan fluorescence, allow for studies of the conformational changes in M-PPases in real-time. The disadvantage of this system over tryptophan fluorescence is the need for the fluorescence dyes, as these are often bulky molecules that are highly flexible. This degree of motion can sometimes drown out smaller changes in conformation. Additionally, it requires that any cysteines not involved in the experiments are removed to prevent unwanted dye binding. In some proteins, this can inactivate the protein, but TmPPase can remain active with only one internal cysteine residue (Harborne,

unpublished), making it possible to carry out these studies. The advantage of this system is that it provides a greater degree of locations to place the dyes. Tryptophan fluorescence experiments rely on the native positions of these residues, if they are not present in locations undergoing conformational changes (such as TmPPase), then no change in fluorescence signal can be detected.

To back up the smFRET work, it is also possible to carry out molecular dynamics simulations to study how the protein might behave in specifically controlled environments. This can be useful for studying the movement of key components of the M-PPase, such as the 5-6 loop, helix 12 or the interactions of residues at the ion gate. This information could be used to investigate key stages of the catalytic cycle, either backing up the data obtained in this study or provide alternative explanations for events during the catalytic cycle.

Finally, the SURFE^{2}R technique could be used to study other M-PPases. This would be particularly useful if TmPPase or another Na^+-PPase could be made to work on the system. The issue with H^+-PPases is that protons can never be removed from the system, and so it is not possible to accurately to study the effects of low protons concentrations on cation pumping. However, it is possible to measure ion pumping in low sodium concentrations, allowing for a much greater range of study on TmPPase, including mutational studies and their effects on sodium pumping. Likewise, the K^+-dependent mutants of PaPPase could be tested on this system to study proton-pumping activity and provide more information to complement the hydrolytic activity assay results.

6.3 Final Conclusions & Significance of the Work to the M-PPase Field

The initial aims of this project were to try to better understand the complete catalytic cycle of M-PPases by providing additional structural and biochemical information. I have achieved this goal through the solving of two new TmPPase crystal structures and the accompanying ion pumping data from the SURFE^{2}R. This information, along with the previously solved structures of TmPPase and VrPPase, and the current work by our collaborators in the Sun lab (Taiwan), have provided key information allowing us to generate an updated model of the catalytic cycle. This provides a clear and logical means by which M-PPases pump H^+/Na^+ and couple this to the hydrolysis of PP_i. There is still much work to be done on M-PPases, as highlighted above, but this model provides invaluable information in the understanding of this family of membrane proteins and the structural information provides a mechanistic insight into how these proteins function. Moreover, this structural information can now be used to help drive drug-design programmes to help target the M-PPases of pathogenic species, such as the malaria-causing *Plasmodium* parasites.

List of References

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., & Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr, 68*(Pt 4), 352-367.

Agmon, N. (1995). The Grotthuss mechanism. *Chemical Physics Letters, 244*, 456-462.

Ahn, S., Milner, A. J., Futterer, K., Konopka, M., Ilias, M., Young, T. W., & White, S. A. (2001). The "open" and "closed" structures of the type-C inorganic pyrophosphatases from Bacillus subtilis and Streptococcus gordonii. *J Mol Biol, 313*(4), 797-811.

Aravind, L., & Koonin, E. V. (1998). The HD domain defines a new superfamily of metal-dependent phosphohydrolases. *Trends Biochem Sci, 23*(12), 469-472.

Asaoka, M., Segami, S., & Maeshima, M. (2014). Identification of the critical residues for the function of vacuolar H(+)-pyrophosphatase by mutational analysis based on the 3D structure. *J Biochem, 156*(6), 333-344.

Ballesteros, J. A., & Weinstein, H. (1995). Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci., 25*, 366-428.

Baltscheffsky, H., Von Stedingk, L. V., Heldt, H. W., & Klingenberg, M. (1966). Inorganic pyrophosphate: formation in bacterial photophosphorylation. *Science, 153*(3740), 1120-1122.

Baykov, A. A., Dubnova, E. B., Bakuleva, N. P., Evtushenko, O. A., Zhen, R. G., & Rea, P. A. (1993). Differential sensitivity of membrane-associated pyrophosphatases to inhibition by diphosphonates and fluoride delineates two classes of enzyme. *FEBS Lett, 327*(2), 199-202.

Baykov, A. A., Malinen, A. M., Luoto, H. H., & Lahti, R. (2013). Pyrophosphate-fueled Na+ and H+ transport in prokaryotes. *Microbiol Mol Biol Rev, 77*(2), 267-276.

Belogurov, G. A., & Lahti, R. (2002). A lysine substitute for K+. A460K mutation eliminates K+ dependence in H+-pyrophosphatase of Carboxydothermus hydrogenoformans. *J Biol Chem, 277*(51), 49651-49654.

Belogurov, G. A., Malinen, A. M., Turkina, M. V., Jalonen, U., Rytkönen, K., Baykov, A. A., & Lahti, R. (2005). Membrane-bound pyrophosphatase of Thermotoga maritima requires sodium for activity. *Biochemistry, 44*, 2088-2096.

Bikadi, Z., Demko, L., & Hazai, E. (2007). Functional and structural characterization of a protein based on analysis of its hydrogen bonding network by hydrogen bonding plot. *Arch Biochem Biophys, 461*(2), 225-234.

Bond, M. W., Chiu, N. Y., & Cooperman, B. S. (1980). Identification of an arginine important for enzymatic activity within the covalent structure of yeast inorganic pyrophosphatase. *Biochemistry, 19*(1), 94-102.

Boyland, E. (1930). Phosphoric esters in alcoholic fermentation: Pyrophosphate in yeast preparations. *Biochem J, 24*(2), 350-354.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem, 72*, 248-254.

Brini, F., Hanin, M., Mezghani, I., Berkowitz, G. A., & Masmoudi, K. (2007). Overexpression of wheat Na+/H+ antiporter TNHX1 and H+-pyrophosphatase TVP1 improve salt- and drought-stress tolerance in Arabidopsis thaliana plants. *J Exp Bot, 58*(2), 301-308.

Bucke, C. (1970). The distribution and properties of alkaline inorganic pyrophosphatase from higher plants. *Phytochemistry, 9*(6), 1303-1309.

Bunick, G., McKenna, G. P., Colton, R., & Voet, D. (1974). The x-ray structure of yeast inorganic pyrophosphatase. Crystal properties. *J Biol Chem, 249*(14), 4647-4649.

Bunick, G., McKenna, P., Scarbrough, F. E., Uberbacher, E. C., & Voet, D. (1978). The X-ray structure of yeast inorganic pyrophosphatase at 5.5 A resolution. *Acta Crystallographica Section B, 34*(11), 3210-3215.

Butler, L. G., & Sperow, J. W. (1977). Multiple roles of metal ions in the reaction catalyzed by yeast inorganic pyrophosphatase. *Bioinorg Chem, 7*(2), 141-150.

Carystinos, G. D., MacDonald, H. R., Monroy, A. F., Dhindsa, R. S., & Poole, R. J. (1995). Vacuolar H(+)-translocating pyrophosphatase is induced by anoxia or chilling in seedlings of rice. *Plant Physiol, 108*(2), 641-649.

Chen, J., Brevet, A., Fromant, M., Lévêque, F., Schmitter, J., Blanquet, S., & Plateau, P. (1990). Pyrophosphatase is essential for growth of *Escherichia coli*. *Journal of Bacteriology, 172*(10), 5686-5689.

Cohn, M. (1958). Phosphate-water exchange reaction catalyzed by inorganic pyrophosphatase of yeast. *J Biol Chem, 230*(1), 369-379.

Cooperman, B. S., & Chiu, N. Y. (1973). Yeast inorganic pyrophosphatase. 3. Active-site mapping by electrophilic reagents and binding measurements. *Biochemistry, 12*(9), 1676-1682.

Cooperman, B. S., Panackal, A., Springs, B., & Hamm, D. J. (1981). Divalent metal ion, inorganic phosphate, and inorganic phosphate analogue binding to yeast inorganic pyrophosphatase. *Biochemistry, 20*(21), 6051-6060.

Dahl, A. C., Chavent, M., & Sansom, M. S. (2012). Bendix: intuitive helix geometry analysis and abstraction. *Bioinformatics, 28*(16), 2193-2194.

Delespaux, V., & de Koning, H. P. (2007). Drugs and drug resistance in African trypanosomiasis. *Drug Resist Updat, 10*(1-2), 30-50.

Docampo, R., de Souza, W., Miranda, K., Rohloff, P., & Moreno, S. N. (2005). Acidocalcisomes - conserved from bacteria to man. *Nat Rev Microbiol, 3*(3), 251-261.

Docampo, R., & Moreno, S. N. (2008). The acidocalcisome as a target for chemotherapeutic agents in protozoan parasites. *Curr Pharm Des, 14*(9), 882-888.

Docampo, R., & Moreno, S. N. (2011). Acidocalcisomes. *Cell Calcium, 50*(2), 113-119.

Docampo, R., Scott, D. A., Vercesi, A. E., & Moreno, S. N. (1995). Intracellular Ca2+ storage in acidocalcisomes of Trypanosoma cruzi. *Biochem J, 310 (Pt 3)*, 1005-1012.

Dondorp, A. M., Nosten, F., Yi, P., Das, D., Phyo, A. P., Tarning, J., Lwin, K. M., Ariey, F., Hanpithakpong, W., Lee, S. J., Ringwald, P., Silamut, K., Imwong, M., Chotivanich, K., Lim, P., Herdman, T., An, S. S., Yeung, S., Singhasivanon, P., Day, N. P. J., Lindegardh, N., Socheat, D., & White, N. J. (2009). Artemisinin Resistance in Plasmodium falciparum Malaria. *The New England journal of medicine, 361*(5), 455-467.

Drozdowicz, Y. M., Lu, Y.-P., Patel, V., Fitz-Gibbon, S., Miller, J. H., & Rea, P. A. (1999). A thermostable vacuolar-type membrane pyrophosphatase from the archaeon Pyrobaculum aerophilum: implications for the origins of pyrophosphate-energized pumps. *FEBS Letters, 460*(3), 505-512.

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 4), 486-501.

Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr, 62*(Pt 1), 72-82.

Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr, 69*(Pt 7), 1204-1214.

Fairhurst, R. M., & Dondorp, A. M. (2016). Artemisinin-Resistant Plasmodium falciparum Malaria. *Microbiol Spectr, 4*(3).

Ferjani, A., Segami, S., Horiguchi, G., Muto, Y., Maeshima, M., & Tsukaya, H. (2011). Keep an eye on PPi: the vacuolar-type H+-pyrophosphatase regulates postgerminative development in Arabidopsis. *Plant Cell, 23*(8), 2895-2908.

Francia, M. E., Wicher, S., Pace, D. A., Sullivan, J., Moreno, S. N., & Arrizabalaga, G. (2011). A Toxoplasma gondii protein with homology to intracellular type Na(+)/H(+) exchangers is important for osmoregulation and invasion. *Exp Cell Res, 317*(10), 1382-1396.

Gao, F., Gao, Q., Duan, X., Yue, G., Yang, A., & Zhang, J. (2006). Cloning of an H+-PPase gene from Thellungiella halophila and its heterologous expression to improve tobacco salt tolerance. *J Exp Bot, 57*(12), 3259-3270.

Garavito, R. M., & Ferguson-Miller, S. (2001). Detergents as tools in membrane biochemistry. *J Biol Chem, 276*(35), 32403-32406.

Garavito, R. M., Picot, D., & Loll, P. J. (1996). Strategies for crystallizing membrane proteins. *J Bioenerg Biomembr, 28*(1), 13-27.

García-Contreras, R., Celis, H., & Romero, I. (2004). Importance of Rhodospirillum rubrum H(+)-Pyrophosphatase under Low-Energy Conditions. *Journal of Bacteriology, 186*(19), 6651-6655.

Gourdon, P., Andersen, J. L., Hein, K. L., Bublitz, M., Pedersen, B. P., Liu, X.-Y., Yatime, L., Nyblom, M., Nielsen, T. T., Olesen, C., Møller, J. V., Nissen, P., & Morth, J. P. (2011). HiLiDe - Systematic approach to membrane protein crystallization in lipid and detergent. *Crystal Growth and Design, 11*(6), 2098-2106.

Harutyunyan, E. H., Kuranova, I. P., Vainshtein, B. K., Höhne, W. E., Lamzin, V. S., Dauter, Z., Teplyakov, A. V., & Wilson, K. S. (1996). X-ray structure of yeast inorganic pyrophosphatase complexed with manganese and phosphate. *Eur J Biochem, 239*(1), 220-228.

Heikinheimo, P., Lehtonen, J., Baykov, A. A., Lahti, R., Cooperman, B. S., & Goldman, A. (1996). the structural basis for pyrophosphatase catalysis. *Structure, 4*(12), 1491-1508.

Heikinheimo, P., Tuominen, V., Ahonen, A. K., Teplyakov, A., Cooperman, B. S., Baykov, A. A., Lahti, R., & Goldman, A. (2001). Toward a quantum-mechanical description of metal-assisted phosphoryl transfer in pyrophosphatase. *Proc Natl Acad Sci U S A, 98*(6), 3121-3126.

Hirono, M., & Maeshima, M. (2009). Functional enhancement by single-residue substitution of Streptomyces coelicolor A3(2) H+-translocating pyrophosphatase. *J Biochem, 146*(5), 617-621.

Hirono, M., Nakanishi, Y., & Maeshima, M. (2007). Essential amino acid residues in the central transmembrane domains and loops for energy coupling of Streptomyces coelicolor A3(2) H+-pyrophosphatase. *Biochim Biophys Acta, 1767*(7), 930-939.

Höhne, W. E., & Heitmann, P. (1974). Tripolyphosphate as a substrate of the inorganic pyrophosphatase from baker's yeast; the role of divalent metal ions. *Acta Biol Med Ger, 33*(1), 1-14.

Hsu, S. H., Hsiao, Y., Liu, P., Lin, S., Luo, Y., & Pan, R. (2009). Purification, characterization, and spectral analyses of histidine-tagged vacuolar H^+-Pyrophosphatase expressed in yeast. *Botanical Studies, 50*, 291-301.

Huang, H., Yury, P., Toro, R., Farelli, J. D., Pandya, C., Almo, S. C., Allen, K. N., & Dunaway-Mariano, D. (2011). Divergence of Structure and Function in the Haloacid Dehalogenase Enzyme Superfamily: Bacteroides thetaiotaomicron BT2127 is an Inorganic Pyrophosphatase(). *Biochemistry, 50*(41), 8937-8949.

Huber, R., Langworthy, T., König, H., Thomm, M., Woese, C., Sleytr, U., & Stetter, K. (1986). Thermotoga maritima sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. *Archives of Microbiology, 144*(4), 324-333.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *J Mol Graph, 14*(1), 33-38, 27-38.

Janson, C. A., Degani, C., & Boyer, P. D. (1979). The formation of enzyme-bound and medium pyrophosphate and the molecular basis of the oxygen exchange reaction of yeast inorganic pyrophosphatase. *J Biol Chem, 254*(10), 3743-3749.

Kabsch, W. (2010). XDS. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 2), 125-132.

Kankare, J., Neal, G. S., Salminen, T., Glumoff, T., Glumhoff, T., Cooperman, B. S., Lahti, R., & Goldman, A. (1994). The structure of E.coli soluble inorganic pyrophosphatase at 2.7 A resolution. *Protein Eng, 7*(7), 823-830.

Karlsson, J. (1975). Membrane-bound potassium and magnesium ion-stimulated inorganic pyrophosphatase from roots and cotyledons of sugar beet (Beta Vulgaris L.). *Biochimica et Biophysica Acta (BBA) - General Subjects, 399*(2), 356-363.

Kay, H. D. (1928). The phosphatases of mammalian tissues: Pyrophosphatase. *Biochem J, 22*(6), 1446-1448.

Kellosalo, J., Kajander, T., Honkanen, R., & Goldman, A. (2013). Crystallization and preliminary X-ray analysis of membrane-bound pyrophosphatases. *Mol Membr Biol, 30*(1), 64-74.

Kellosalo, J., Kajander, T., Kogan, K., Pokharel, K., & Goldman, A. (2012). The structure and catalytic cycle of a sodium-pumping pyrophosphatase. *Science, 337*(6093), 473-476.

Kellosalo, J., Kajander, T., Palmgren, M. G., Lopéz-Marqués, R. L., & Goldman, A. (2011). Heterologous expression and purification of membrane-bound pyrophosphatases. *Protein Expr Purif, 79*(1), 25-34.

Kim, E. J., Zhen, R. G., & Rea, P. A. (1994). Heterologous expression of plant vacuolar pyrophosphatase in yeast demonstrates sufficiency of the substrate-binding subunit for proton transport. *Proc Natl Acad Sci U S A, 91*(13), 6128-6132.

Kornberg, A. (1962). On the metabolic significance of phosphorolytic and pyrophosphorolytic reactions. In M. Kasha & B. Pullman (Eds.), *Horizons in Biochemistry* (pp. 251-264). New York: Academic Press.

Krogstad, D. J., Gluzman, I. Y., Kyle, D. E., Oduola, A. M., Martin, S. K., Milhous, W. K., & Schlesinger, P. H. (1987). Efflux of chloroquine from Plasmodium falciparum: mechanism of chloroquine resistance. *Science, 238*(4831), 1283-1285.

Kunitz, M. (1951). Crystalline inorganic pyrophosphatase isolated from baker's yeast. *The Journal of General Physiology*, 423-450.

Lee, H. S., Cho, Y., Kim, Y. J., Lho, T. O., Cha, S. S., Lee, J. H., & Kang, S. G. (2009). A novel inorganic pyrophosphatase in Thermococcus onnurineus NA1. *FEMS Microbiol Lett, 300*(1), 68-74.

Lemercier, G., Dutoya, S., Luo, S., Ruiz, F. A., Rodrigues, C. O., Baltz, T., Docampo, R., & Bakalara, N. (2002). A vacuolar-type H+-pyrophosphatase governs maintenance of functional acidocalcisomes and growth of the insect and mammalian forms of Trypanosoma brucei. *J Biol Chem, 277*(40), 37369-37376.

Lemercier, G., Espiau, B., Ruiz, F. A., Vieira, M., Luo, S., Baltz, T., Docampo, R., & Bakalara, N. (2004). A pyrophosphatase regulating polyphosphate metabolism in acidocalcisomes is essential for Trypanosoma brucei virulence in mice. *J Biol Chem, 279*(5), 3420-3425.

Li, J., Yang, H., Peer, W. A., Richter, G., Blakeslee, J., Bandyopadhyay, A., Titapiwantakun, B., Undurraga, S., Khodakovskaya, M., Richards, E. L., Krizek, B., Murphy, A. S., Gilroy, S., & Gaxiola, R. (2005). Arabidopsis H+-PPase AVP1 regulates auxin-mediated organ development. *Science, 310*(5745), 121-125.

Li, K. M., Wilkinson, C., Kellosalo, J., Tsai, J. Y., Kajander, T., Jeuken, L. J., Sun, Y. J., & Goldman, A. (2016). Membrane pyrophosphatases from Thermotoga maritima and Vigna radiata suggest a conserved coupling mechanism. *Nat Commun, 7*, 13596.

Lim, L., & McFadden, G. I. (2010). The evolution, metabolism and functions of the apicoplast. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*(1541), 749-763.

Lin, S. M., Tsai, J. Y., Hsiao, C. D., Huang, Y. T., Chiu, C. L., Liu, M. H., Tung, J. Y., Liu, T. H., Pan, R. L., & Sun, Y. J. (2012). Crystal structure of a membrane-embedded H+-translocating pyrophosphatase. *Nature, 484*(7394), 399-403.

Liu, J., Pace, D., Dou, Z., King, T. P., Guidot, D., Li, Z.-H., Carruthers, V. B., & Moreno, S. N. J. (2014). A Vacuolar-H(+)-Pyrophosphatase (TgVP1) is Required for Microneme Secretion, Host Cell Invasion, and Extracellular Survival of Toxoplasma gondii. *Molecular microbiology, 93*(4), 698-712.

López-Marqués, R. L., Pérez-Castiñeira, J. R., Buch-Pedersen, M. J., Marco, S., Rigaud, J. L., Palmgren, M. G., & Serrano, A. (2005). Large-scale purification of the proton pumping pyrophosphatase from Thermotoga maritima: a "Hot-Solve" method for isolation of recombinant thermophilic membrane proteins. *Biochim Biophys Acta, 1716*(1), 69-76.

López-Marqués, R. L., Pérez-Castiñeira, J. R., Losada, M., & Serrano, A. (2004). Differential regulation of soluble and membrane-bound inorganic

pyrophosphatases in the photosynthetic bacterium Rhodospirillum rubrum provides insights into pyrophosphate-based stress bioenergetics. *J Bacteriol, 186*(16), 5418-5426.

Lundin, M., Baltscheffsky, H., & Ronne, H. (1991). Yeast PPA2 gene encodes a mitochondrial inorganic pyrophosphatase that is essential for mitochondrial function. *J Biol Chem, 266*(19), 12168-12172.

Luoto, H. H., Baykov, A. A., Lahti, R., & Malinen, A. M. (2013). Membrane-integral pyrophosphatase subfamily capable of translocating both Na+ and H+. *PNAS, 110*(4), 1255-1260.

Luoto, H. H., Belogurov, G. A., Baykov, A. A., Lahti, R., & Malinen, A. M. (2011). Na+-translocating membrane pyrophosphatases are widespread in the microbial world and evolutionarily precede H+-translocating pyrophosphatases. *J Biol Chem, 286*(24), 21633-21642.

Maeshima, M. (1991). H+-translocating inorganic pyrophosphatase of plant vacuoles Inhibition by Ca2+, stabilization by Mg2+ and immunological comparison with other inorganic pyrophosphatases. *European Journal of Biochemistry, 196*(1), 11-17.

Malinen, A. M., Belogurov, G. A., Baykov, A. A., & Lahti, R. (2007). Na+-pyrophosphatase: a novel primary sodium pump. *Biochemistry, 46*, 8872-8878.

Martinez, R., Wang, Y., Benaim, G., Benchimol, M., de Souza, W., Scott, D. A., & Docampo, R. (2002). A proton pumping pyrophosphatase in the golgi apparatus and plasma membrane vesicles of Trypanosoma cruzi. *Molecular & Biochemical Parasitology, 120*, 205-213.

McFadden, G. I. (2000). Mergers and acquisitions: malaria and the great chloroplast heist. *Genome Biol, 1*(4), REVIEWS1026.

McIntosh, M. T., Drozdowicz, Y. M., Laroiya, K., Rea, P. A., & Vaidya, A. B. (2001). Two classes of plant-like vacuolar-type H+-pyrophosphatases in malaria parasites. *Molecular & Biochemical Parasitology, 114*, 183-195.

Merckel, M. C., Fabrichniy, I. P., Salminen, A., Kalkkinen, N., Baykov, A. A., Lahti, R., & Goldman, A. (2001). Crystal structure of Streptococcus mutans pyrophosphatase: a new fold for an old mechanism. *Structure, 9*(4), 289-297.

Moe, O. A., & Butler, L. G. (1972). Yeast inorganic pyrophosphatase. II. Kinetics of Mg 2+ activation. *J Biol Chem, 247*(22), 7308-7314.

Moreno, S. N., & Zhong, L. (1996). Acidocalcisomes in Toxoplasma gondii tachyzoites. *Biochem. J., 313*, 655-659.

Moyle, J., Mitchell, R., & Mitchell, P. (1972). Proton translocating pyrophosphatase of *Rhodpirillum rubrum*. *FEBS Letters, 23*(2), 233-236.

Mulkidjanian, A. Y., Galperin, M. Y., Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2008). Evolutionary primacy of sodium bioenergetics. *Biol Direct, 3*, 13.

Nakanishi, Y., Yabe, I., & Maeshima, M. (2003). Patch clamp analysis of a H+ pump heterologously expressed in giant yeast vacuoles. *J Biochem, 134*(4), 615-623.

Nayal, M., & Di Cera, E. (1996). Valence screening of water in protein crystals reveals potential Na+ binding sites. *J Mol Biol, 256*(2), 228-234.

Negi, T., Samejima, T., & Irie, M. (1972). Studies on the tryptophan residues of yeast inorganic pyrophosphatase in relation to the enzymatic activity. *J Biochem, 71*(1), 29-37.

Nordbo, E., Luoto, H. H., Baykov, A. A., Lahti, R., & Malinen, A. M. (2016). Two independent evolutionary routes to Na+/H+ cotransport function in membrane pyrophosphatases. *Biochem J, 473*(19), 3099-3111.

Nyrén, P., & Strid, Å. (1991). Hypothesis: the physiological role of the membrane-bound proton-translocating pyrophosphatase in some phototrophic bacteria. *FEMS Microbiology Letters, 77*(2-3), 265-270.

Oldham, M. L., Hite, R. K., Steffen, A. M., Damko, E., Li, Z., Walz, T., & Chen, J. (2016). A mechanism of viral immune evasion revealed by cryo-EM analysis of the TAP transporter. *Nature, 529*(7587), 537-540.

Oliver, R. C., Lipfert, J., Fox, D. A., Lo, R. H., Doniach, S., & Columbus, L. (2013). Dependence of micelle size and shape on detergent alkyl chain length and head group. *PLoS One, 8*(5), e62488.

Parfenyev, A. N., Salminen, A., Halonen, P., Hachimori, A., Baykov, A. A., & Lahti, R. (2001). Quaternary structure and metal ion requirement of family II pyrophosphatases from Bacillus subtilis, Streptococcus gordonii, and Streptococcus mutans. *J Biol Chem, 276*(27), 24511-24518.

Pasapula, V., Shen, G., Kuppu, S., Paez-Valencia, J., Mendoza, M., Hou, P., Chen, J., Qiu, X., Zhu, L., Zhang, X., Auld, D., Blumwald, E., Zhang, H., Gaxiola, R., & Payton, P. (2011). Expression of an Arabidopsis vacuolar H+-pyrophosphatase gene (AVP1) in cotton improves drought- and salt tolerance and increases fibre yield in the field conditions. *Plant Biotechnol J, 9*(1), 88-99.

Pérez-Castiñeira, J. R., Alvar, J., Ruiz-Pérez, L. M., & Serrano, A. (2002). Evidence for a wide occurrence of proton-translocating pyrophosphatase genes in parasitic and free-living protozoa. *biochemical and Biophysical Research Communications, 294*, 567-573.

Pérez-Castiñeira, J. R., López-Marqués, R. L., Losada, M., & Serrano, A. (2001). A thermostable K+-stimulated vacuolar-type pyrophosphatase from the hyperthermophilic bacterium Thermotoga maritima. *FEBS Letters, 496*, 6-11.

Rapoport, T. A., Höhne, W. E., Heitmann, P., & Rapoport, S. (1973). Binding of ligands to the inorganic pyrophosphatase of bakers' yeast. *Eur J Biochem, 33*(2), 341-347.

Rea, P. A., Kim, Y., Sarafian, V., Poole, R. J., Davies, J. M., & Sanders, D. (1992). Vacuolar H(+)-translocating pyrophosphatases: a new category of ion translocase. *Trends Biochem Sci, 17*(9), 348-353.

Rodrigues, C. O., Scott, D. A., Bailey, B. N., de Souza, W., Benchimol, M., Moreno, B., Urbina, J. A., Oldfield, E., & Moreno, S. N. (2000). Vacuolar proton pyrophosphatase activity and pyrophosphate (PPi) in Toxoplasma gondii as possible chemotherapeutic targets. *Biochem. J., 349*, 737-745.

Rodrigues, C. O., Scott, D. A., & Docampo, R. (1999). Presence of a vacuolar H+-pyrophosphatase in promastigotes of Leishmania donovani and its localisation to a different compartment from the vacuolar H+-ATPase. *Biochem. J., 340*, 759-766.

Rohloff, P., Montalvetti, A., & Docampo, R. (2004). Acidocalcisomes and the contractile vacuole complex are involved in osmoregulation in Trypanosoma cruzi. *J Biol Chem, 279*(50), 52270-52281.

Salminen, T., Käpylä, J., Heikinheimo, P., Kankare, J., Goldman, A., Heinonen, J., Baykov, A. A., Cooperman, B. S., & Lahti, R. (1995). Structure and function analysis of Escherichia coli inorganic pyrophosphatase: is a hydroxide ion the key to catalysis? *Biochemistry, 34*(3), 782-791.

Scott, D. A., de Souza, W., Benchimol, M., Zhong, L., Lu, H. G., Moreno, S. N. J., & Docampo, R. (1998). Presence of a Plant-like Proton-pumping Pyrophosphatase in Acidocalcisomes of Trypanosoma cruzi. *Journal of Biological Chemistry, 273*(34), 22151-22158.

Seufferheld, M., Vieira, M. C., Ruiz, F. A., Rodrigues, C. O., Moreno, S. N., & Docampo, R. (2003). Identification of organelles in bacteria similar to acidocalcisomes of unicellular eukaryotes. *J Biol Chem, 278*(32), 29971-29978.

Shapiro, T. A., & Englund, P. T. (1995). The structure and replication of kinetoplast DNA. *Annu Rev Microbiol, 49*, 117-143.

Shen-Hsing, H., Yi-Yuong, H., Pei-Feng, L., Shih-Ming, L., Yue-Yu, L., & Rong-Long, P. (2009). Purification, characterization, and spectral analyses of histidine-tagged vacuolar H+-pyrophosphatase expressed in yeast. *Botanical Studies, 50*, 291-301.

Shintani, T., Uchiumi, T., Yonezawa, T., Salminen, A., Baykov, A. A., Lahti, R., & Hachimori, A. (1998). Cloning and expression of a unique inorganic pyrophosphatase from *Bacillus subtilis*: evidence for a new family of enzymes. *FEBS Letters, 439*, 263-266.

Springs, B., Welsh, K. M., & Cooperman, B. S. (1981). Thermodynamics, kinetics, and mechanism in yeast inorganic pyrophosphatase catalysis of inorganic pyrophosphate: inorganic phosphate equilibration. *Biochemistry, 20*(22), 6384-6391.

Tsai, J. Y., Kellosalo, J., Sun, Y. J., & Goldman, A. (2014). Proton/sodium pumping pyrophosphatases: the last of the primary ion pumps. *Curr Opin Struct Biol, 27*, 38-47.

Venancio, J. B., Catunda, M. G., Ogliari, J., Rima, J. A., Okorokova-Facanha, A. L., Okorokov, L. A., & Facanha, A. R. (2014). A vacuolar H(+)-pyrophosphatase differential activation and energy coupling integrate the responses of weeds and crops to drought stress. *Biochim Biophys Acta, 1840*(6), 1987-1992.

Völkl, P., Huber, R., Drobner, E., Rachel, R., Burggraf, S., Trincone, A., & Stetter, K. O. (1993). Pyrobaculum aerophilum sp. nov., a novel nitrate-reducing hyperthermophilic archaeum. *Applied and Environmental Microbiology, 59*(9), 2918-2926.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., & Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr, 67*(Pt 4), 235-242.

Yano, Y., Negi, T., & Irie, M. (1973). Carboxamidomethylation of yeast inorganic pyrophosphatase. *J Biochem, 74*(1), 67-76.

Yuvaniyama, J., Chitnumsub, P., Kamchonwongpaisan, S., Vanichtanankul, J., Sirawaraporn, W., Taylor, P., Walkinshaw, M. D., & Yuthavong, Y. (2003). Insights into antifolate resistance from malarial DHFR-TS structures. *Nat Struct Biol, 10*(5), 357-365.

Zheng, H., Cooper, D. R., Porebski, P. J., Shabalin, I. G., Handing, K. B., & Minor, W. (2017). CheckMyMetal: a macromolecular metal-binding validation tool. *Acta Crystallogr D Struct Biol, 73*(Pt 3), 223-233.

Zyryanov, A. B., Vener, A. V., Salminen, A., Goldman, A., Lahti, R., & Baykov, A. A. (2004). Rates of Elementary Catalytic Steps for Different Metal Forms of the Family II Pyrophosphatase from Streptococcus gordonii. *Biochemistry, 43*(4), 1065-1074.