Towards more rational approaches of membrane protein stabilisation and novel structures of membrane-bound pyrophosphatase

Jannik Strauss

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

School of Molecular and Cellular Biology

&

The Astbury Centre of Structural Molecular Biology

December 2021

Intellectual Property and Publication Statement

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors of this work has been explicitly indicated below. The candidate confirms that appropriate credit had been given within the thesis where reference has been made to the work of others.

Components of this thesis contain work from jointly authored publications. Co-first authorship is indicated by an asterisk.

Chapter 1 contains figures adapted from following research articles and reviews:

(I) Ehud M. Landau, Jürg P. Rosenbusch. Lipidic cubic phases: A novel concept for the crystallization of membrane proteins. PNAS. 1996 Dec 10, 93 (25) 14532-14535.

An adaptation of Figure 1 from (I) was used in the introduction.

Permission of redistribution: Permission to use adapted figures from (I) is not required if the article is published under the Exclusive PNAS License to Publish and only used for non-commercial or educational purposes, as done here. Appropriate credit must be given and following copyright statement must be included: Copyright (1996) National Academy of Sciences.

Chapter 2 contains methodologies and written content adapted from following jointly authored articles:

(II) **Strauss J**, Wilkinson C, Vidilaseris K, Harborne SPD, Goldman A. A Simple Strategy to Determine the Dependence of Membrane-Bound Pyrophosphatases on K+ as a Cofactor. In: Allen KN, editor. Methods in Enzymology. Academic Press; 2018. p. 131–56. (Phosphatases; vol. 607).

(III) Harborne SPD*, **Strauss J***, Boakes JC*, Wright DL, Henderson JG, Boivineau J, Jaakola VP, Goldman A. IMPROvER: the Integral Membrane Protein Stability Selector. Sci Rep. 2020 Sep 16;10(1):15165

(IV) Cecchetti C*, **Strauss J***, Stohrer C, Naylor C, Pryor E, Hobbs J, Tanley J, Goldman A, Byrne B. A novel high-throughput screen for identifying lipids that stabilise membrane proteins in detergent based solution. PLOS ONE. 2021 Jul 12;16(7):e0254118.

Permission of redistribution: Permission to use adapted tables and figures from (II) was obtained from Elsevier under licence number 5113091129587. Articles (III) and (IV) are published under a Creative Commons Attribution International

Chapter 3 contains data, analysis and written content adapted from following co-first authored articles. Detailed information about the candidate's involvement is given below:

(V) Harborne SPD*, **Strauss J***, Boakes JC*, Wright DL, Henderson JG, Boivineau J, Jaakola VP, Goldman A. IMPROvER: the Integral Membrane Protein Stability Selector. Sci Rep. 2020 Sep 16;10(1):15165.

Author Contributions: S.P.D.H. and A.G. conceived the study and planned the experiments. S.P.D.H. wrote the computational pipeline. J.S., J.C.B. and J.G.H. carried out literature searches for critical residues of *Cl*-PPase, hENT1 and hPTH1R, respectively. J.S. carried out mutagenesis and stability testing of *Cl*-PPase and developed a rapid screening strategy. S.P.D.H., D.L.W., J.C.B., J.B. and V.-P.J. carried out mutagenesis and stability testing of hENT1. J.G.H. carried out stability testing of hPTH1R. S.P.D.H., J.S. and A.G. wrote the manuscript. All authors assisted manuscript revision and editing.

Adaptations of Figure 1, Figure 3, Figure S1-5, Table 1 and Table S2 from V are used in this thesis alongside written content. The candidate devised and wrote the manuscript sections describing work related to *Cl*-PPase and prepared corresponding tables and figures under guidance of Dr. Steven Harborne (Peak Proteins) and Prof. Adrian Goldman (University of Helsinki).

(VI) Cecchetti C*, **Strauss J***, Stohrer C, Naylor C, Pryor E, Hobbs J, Tanley J, Goldman A, Byrne B. A novel high-throughput screen for identifying lipids that stabilise membrane proteins in detergent based solution. PLOS ONE. 2021 Jul 12;16(7):e0254118.

Author Contributions: CC, JS, JH, BB and AG conceived the study and planned the experiments. CC, JS, CN, EP and BB carried out literature research on lipids and designed the lipid screen. CC and JS produced the first batch of the Lipid Screen. JH and ST supervised the work in industry. CC, JS, CS expressed and purified test proteins ($A_{2A}R$, UapA, *Tm*-PPase) for stability testing and data analysis carried out by JS. CC, CS, AG and BB wrote the manuscript. All authors assisted in manuscript revision and editing.

Adaptations of Figure 1-3, Figure S1 and Table 1 from V are used in this thesis alongside written content. The candidate devised and wrote large parts of the manuscript including the introduction, general methods/results/discussion and all

sections describing work related to *Tm*-PPase under guidance of Prof. Bernadette Byrne (Imperial College London) and Prof. Adrian Goldman. The candidate made all figures and tables.

Chapter 4 contains data from following PhD project of former group member Dr. Craig Wilkinson. Detailed information about the candidate's involvement is given below.

(VII) Wilkinson C. Understanding the Catalytic Cycle of Membrane Pyrophosphatases Through Structural and Functional Studies. PhD thesis. University of Leeds; 2017 Oct. http://etheses.whiterose.ac.uk/19131/

An adaption of Table 4.5 from VII is used in this thesis and written consent is given below. The candidate was not involved in any way in the abovementioned PhD project. The candidate used established expression and purification protocols to produce wildtype and variant *Pa*-PPase for functional and structural characterisation as well as X-ray diffraction data of *Pa*-PPase crystals obtained by Dr. Craig Wilkinson. The candidate received help for expressing and purifying *Pa*-PPase variants by Dr. James Hillier (University of Leeds).

Chapter 5 is partly based on previous unpublished work by former group member Dr. Nita Shah.

The candidate used previously established expression and purification protocols to produce wildtype and variant *Cl*-PPase for detergent-screening and crystallisation trials. Please contact Prof. Adrian Goldman to obtain relevant data of previous work undertook by Dr. Nita Shah. Mass spectrometry data was obtained by Dr. Rachel George from the Mass Spectrometry Facility at the University of Leeds upon supply of purified proteins.

Chapter 6 does not contain data that is part of previous studies.

The candidate received help for the expression and purification of wild-type *Tm*-PPase from Dr. Keni Vidilaseris (University of Helsinki), Orquidea De Castro Ribeiro (University of Helsinki) and Dr. Bernadette Gehl-Väisänen (University of Aalto). Serial X-ray diffraction data obtained at beamline P14-II (DESY) was indexed by Dr. David von Stetten (EMBL Hamburg). Moreover, an adapted image of Figure 3 of VIII was used for illustration purposes from the following publication:

(VIII) Mehrabi P, Muller-Werkmeister HM, Leimkohl J-P., Schikora H, Ninkovic J, Krivokuca S, Andricek L, Epp SW, Sherrell D, Owen RL, Pearson AR, Tellkamp F., Schulz EC and Miller RJD. The HARE chip for efficient time-resolved serial synchrotron crystallography. J Synchrotron Rad. 2020. 27, 360-370.

Chapter 7 summarises Chapter 3-6 and contains data and analysis referred to above.

Acknowledgements

First of all, I would like to thank my supervisors Prof. Adrian Goldman, Dr. Christos Pliotas and Dr. Antreas Kalli from the University of Leeds and Prof. Bernadette Byrne from Imperial College London for their support and guidance over the past few years. Their input did not only help me to navigate the world of membrane proteins, but also shaped me as a scientist and ensured the successful completion of my PhD project.

I have been extraordinarily fortunate to join an amazing team here at the University of Leeds (and in Helsinki) that made me feel most welcome and was always there to help. A big thank you goes to our former post-docs Dr. Steven Harborne, Dr. Maren Thomsen and Dr. Julie Heggelund who took the time to answer all my questions and dispel self-doubts when a purification failed once again. I greatly appreciate that you were always willing to give advice, even when you had already left our group. The same applies to my predecessor Dr. Craig Wilkinson who helped me countless times with all aspects of M-PPase work throughout the years, spent hours proof-reading my texts and taught me graphology, an unnecessarily important skill. I would also like to thank everyone from RAMP and Astbury level 6 for the many thoughtful discussions and sharing of expertise.

This is the single most challenging and biggest thing I have done in my life but, luckily, I did not have to see it through all alone. Thanks to Alex for the meme artwork, to Sophie for pub nights, to Julie for space cats, to Jess for basically solving all our lab problems, to Claudia for saving my finger when Andreas could not, to Andreas for trying to save my finger, to Jamie for letting me win at science, to Jack for almost beating Jamie at climbing, to Craig for the most entertaining feedback, to Brendan for providing Havana Club, to El for keeping me sane during lockdown and to Maria for poppers and all the adventures.

And to all my friends back home: Vielen Dank für eure Unterstützung und euren Glauben an mich! Ein besonderer Dank geht an Tim und Valeria für die Sicherstellung meines leiblichen Wohls in besonders schweren Zeiten und an Finja, die mich immer ermutigt hat über meinen Schatten zu springen und meine beruflichen Ziele zu verfolgen.

Zuletzt möchte ich meinen größten Dank meiner Freundin und Familie aussprechen, euch widme ich diese Arbeit. Alex, dir danke ich von Herzen für deinen Zuspruch, dein Verständnis und für die vielen kleinen pick-me-ups auf der Zielgeraden. Sabine, Egbert, Luca, Alice und Fritz, ohne eure uneingeschränkte Unterstützung bei all meinen Entscheidungen bis hier hin wäre mein Werdegang so nicht möglich gewesen. Ihr habt einen großen Anteil an meinem Erfolg und dafür bin ich euch sehr dankbar.

Abstract

Membrane proteins have a range of crucial biological functions and are targeted by most prescribed drugs despite lagging behind soluble proteins when it comes to their biochemical and biophysical characterisation. A major bottleneck in membrane protein research is protein instability upon extraction. Protein stabilisation strategies are typically expensive and labour-intensive. Therefore, I contributed to the development and evaluation of two new general-purpose tools, both designed for the streamlined and rational stabilisation of membrane proteins. The first tool, the integral membrane protein stability selector (IMPROvER), predicts stabilising point-mutations in membrane proteins using three individual approaches with additive prediction power. The second tool, a novel pre-prepared and easy-to-use screen for the high-throughput identification of stabilising lipids, facilitates the structural and functional analysis of stable and physiologically relevant protein sample. Both tools were successfully employed to stabilise a range of membrane proteins with different folds, topologies and modes of action at significantly reduced cost and work effort.

Moreover, engineered or natively thermostable membrane-bound pyrophosphatases (M-PPase), were studied in more detail using conventional and time-resolved X-ray crystallography. Based on structural data obtained on a pyrophosphate-energised K^+-independent H^+-pump, I derived an updated model of ion-selectivity that is centred on a glutamate-serine interplay at the ion-gate. This is the first model that explains ion-selectivity in all M-PPase subclasses when considering functional asymmetry. Indeed, complementary time-resolved structural studies of a K^+-dependent Na^+-pump revealed asymmetric substrate binding to M-PPase active sites. These findings give valuable mechanistic insights into key processes of M-PPase biochemistry, which are of upmost importance for structure-guided drug discovery. Ultimately, tweaking M-PPase function has the potential to address existing and emerging challenges to human health and global food security as M-PPases play a vital role in the stress resistance of pathogens or salt and drought resistance in plants.

Table of Contents

List of Figures

Appendix

List of Tables

Appendix

List of Abbreviations

Table 1.0: List of Abbreviations

Abbreviation	Composition
3D	Three-dimensional
A_{280}	Light absorbance at 280 nm
$A_{2A}R$	Adenosine A_{2A} receptor
AMDP	Amino-methylene-diphosphonate
ATC	N-[(2-amino-6-benzothiazolyl)methyl]-1H-indole-2-carboxamide
ATP	Adenosine triphosphate
B&S	Ballesteros & Weinstein
BSA	Bovine serum albumin
CCCP	Carbonyl cyanide m-chlorophenyl hydrazone
CI	Confidence interval
CMC	Critical micelle concentration
Cryo-EM	Cryogenic electron microscopy
DESY	Deutsches Elektronen-Synchrotron
DNA	Deoxyribonucleic acid
Detergents	
C6	Cymal 6
DDM	n-Dodecyl-β-D-Maltopyranoside
DM	n-Decyl-β-D-Maltopyranoside
OGNG	Octyl Glucose Neopentyl Glycol
LDAO	Lauryl dimethylamine oxide
NG	n-Nonyl-α-D-Glucopyranoside
NM	n-Nonyl-β-D-Maltopyranoside
OG	n-Octyl-α-D-Glucopyranoside
OM	n-Octyl-β-D-Maltopyranoside
DiDiMa	Difference distance matrix
DSF	Differential scanning fluorometry
DTT	Dithiothreitol
EDTA	Ethylenediamine tetra acetic acid
F_{330}	Fluorescence signal at 300 nm
F_{350}	Fluorescence signal at 300 nm
FPLC	Fast Protein Liquid Chromatography
GFP	Green fluorescent protein
GPCRs	G-protein coupled receptors
HARE	Hit-and-return
HEDP	1 hydroxyethylidene-1,1-diphosphonate
hENT	Human equilibrative nucleoside transporter 1
HiLiDe	High-lipid detergent crystallisation
HPLC	High-pressure liquid chromatography
hrd	Homologous recombination domain
HRP	Horseradish peroxidase
HRV3C	MBP and a human rhinovirus 3C protease
IC_{50}	Inhibitory concentration 50

IDP	Imidodiphosphate
IMAC	**Immobilised metal affinity chromatography**
BE	buffer exchanged elution fraction
E	elution fraction
FT	flow-through
R	resin
W	wash fraction
IPTG	Isopropyl-β-D-thiogalactoside
ITN	Innovative training network
k_{cat}	Catalytic turnover number
K_d	Dissociation constant
K_m	Microscopic Michaelis-Menten
LAMA	Liquid application method for time-resolved analysis
LB	Lysogeny broth
LC	Liquid chromatography
LCP	Lipidic cubic phase
Lipids	
10:PA	1,2-Didecanoyl-*sn*-Glycero-3-Phosphate (sodium salt)
14:0 CL	1',3'-bis[1,2-dimyristoyl-*sn*-glycero-3-phospho]-glycerol (sodium salt)
16:0-18:1 PA (POPA)	1-Palmitoyl-2-Oleoyl-*sn*-Glycero-3-Phosphate (sodium salt)
18:1 CL	1',3'-bis[1,2-dioleoyl-*sn*-glycero-3-phospho]-glycerol (sodium salt)
CHL	Cholesterol
CHS	Cholesteryl hemisuccinate Tris salt
CL	Cardiolipin
DMPC	1,2-Dimyristoyl-*sn*-Glycero-3-Phosphocholine
DOPC	1,2-Dioleoyl-*sn*-Glycero-3-Phosphocholine
DOPE	1,2-Dioleoyl-*sn*-Glycero-3-Phosphoethanolamine
DOPS	1,2-Dioleoyl-*sn*-Glycero-3-Phospho-L-Serine
DPPC	1,2-Dihexadecanoyl-*sn*-Glycero-3-Phosphocholine
DPPE	1,2-Dipalmitoyl-*sn*-Glycero-3-Phosphoethanolamine
DPPG	1,2-Dipalmitoyl-*sn*-Glycero-3-Phospho-1'-*rac*-Glycerol
EPL	*E. coli* polar lipid extract
LMPC	1-Myristoyl-2-Hydroxy-*sn*-Glycero-3-Phosphocholine
MAG	Monoacylglycerol
7.7MAG	Monomyristolein
9.9MAG	Monoolein
PA	phosphatidic acid
PBL	Porcine polar brain lipid extract
PC	Phosphatidylcholine
PE	Phosphatidylethanolamine
PG	Phosphoglyceride
PI	Phosphatidyl inositol
POPC	1-Palmitoyl-2-Oleoyl-*sn*-Glycero-3-Phosphocholine
POPE	1-Palmitoyl-2-Oleoyl-*sn*-Glycero-3-Phosphoethanolamine
POPG	1-Palmitoyl-2-Oleoyl-*sn*-Glycero-3-Phosphoglycerol
POPS	1-Palmitoyl-2-Oleoyl-*sn*-Glycero-3-Phospho-L-Serine
PS	Phosphatidylserine

SL	Sphingolipid
SM	Egg sphingomyelin (chicken)
MBP	Maltose-binding protein
MD	Molecular dynamics
MEDP	Methylene-diphosphonate
MEM	Resuspended membrane sample
M-PPase	**Membrane-bound pyrophosphatase**
Am-PPase	from *Akkermansia municiphila*
AVP1	from *Arabidopsis thaliana* (isoform 1)
Bv-PPase	from *Bacteroides vulgatus*
Cc-PPase	from *Clostridium lentocellum*
Cl-PPase	from *Clostridium leptum*
Chl-PPase	from *Chlorobium limicola*
Fj-PPase	from *Flavobacterium johnsoniae*
H^+-PPase	H^+-pumping M-PPase
Na^+-PPase	Na^+-pumping M-PPase
Na^+,H^+-PPase	Na^+,H^+-pumping M-PPase
Pa-PPase	from *Pyrobaculum aerophilum*
Po-PPase	from *Prevoltella oralis*
Rb-PPase	from *Rhodospirillum rubrum*
Sc-PPase	from *Streptomyces coelicolor*
Tm-PPase	from *Thermotoga maritima*
Vr-PPase	from *Vigna radiata*
MS	Mass spectrometry
MWCO	Molecular weight cut off
Na_vAb	Voltage-gated sodium channel
NCS	Non-crystallographic symmetry
NMR	Nuclear magnetic resonance
NTA	Nitrilotriacetic-acid
OD	Optical density
Organisms	
A. municiphila	*Akkermansia municiphila*
A. thaliana	*Arabidopsis thaliana* (isoform 1)
B. subtilis	*Bacillus subtilis*
B. vulgatus	*Bacteroides vulgatus*
C. lentocellum	*Clostridium lentocellum*
C. leptum	*Clostridium leptum*
C. limicola	*Chlorobium limicola*
D. hafniensee	*Desulfitobacterium hafniensee*
E. coli	*Escherichia coli*
F. johnsoniae	*Flavobacterium johnsoniae*
P. falciparum	*Plasmodium falciparum*
P. aerophilum	*Pyrobaculum aerophilum*
P. oralis	*Prevoltella oralis*
R. rubrum	*Rhodospirillum rubrum*
S. cerevisiae	*Saccharomyces cerevisiae*
S. gordonii	*Streptococcus gordonii*

S. mutans	*Streptococcus mutans*
S. coelicolor	*Streptomyces coelicolor*
T. maritima	*Thermotoga maritima*
T. brucei	*Trypanosoma brucei*
V. radiata	*Vigna radiata*
PAGE	Polyacrylamide gel electrophoresis
PPase	Pyrophosphatase
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
PELDOR	Pulsed Electron-Electron Double Resonance
PES	Polyether sulfone
P_i	Orthophosphate
pI	Isoelectric point
PMSF	Phenylmethylsulfonylfluoride
PPases	Pyrophosphatases
PP_i	Pyrophosphate, Diphosphate
PTH_1R	Human parathyroid hormone 1 receptor
PVDF	Polyvinylidene fluoride
R^2	Coefficient of determination
RAMP	Rationalising membrane protein crystallisation
RFU	Relative fluorescence units
Rmsd	Root mean square deviation
RNA	Ribonucleic acid
SCD-L/U	Synthetic complete dropout minus leucine/uracil
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SEM	Standard error of the mean
Sf	Superfolder
SMA	Styrene-maleic acid
SMALP	Styrene-maleic acid lipid particles
SOC	Super optimal broth catabolite repression
SOL	Detergent-solubilised crude membrane extract
S-PPase	Soluble pyrophosphatase
TAE	Tris-acetate buffer
TB	Terrific broth
TBF	Transformation buffer
TBS	Tris-buffered saline
TBS-T	Tris-buffered saline + tween 20
TEV	Tobacco Etch Virus
TLS	Translation/ Libration/ Screw
T_m	Apparent melting temperature
UapA	eukaryotic purine/H^+ symporter
VdW	Van der Waals
YPD	Yeast extract Peptone dextrose media
ZM	ZM241385
$\bar{x}$	Average

Chapter 1. Introduction

1.1. A brief introduction to membrane proteins

Membrane proteins are a class of proteins that localise at, or interact with, biological membranes as implied by their name. This class is subdivided into proteins that are permanently (integral) or only temporarily (peripheral) associated to the membrane. In this work, I focus exclusively on integral membrane proteins, more precisely, on proteins that span the membrane multiple times, which are called poly/bitopic- or transmembrane proteins. Their surface is mostly hydrophobic to allow a stable incorporation in the lipid bilayer. In contrast, soluble proteins mainly expose hydrophilic residues on their surface and are therefore found in aqueous solutions such as the cytoplasm or periplasm of cells. Peripheral membrane proteins are also largely water soluble as they adhere to the membrane by binding to their integral counterparts or only peripherally penetrate the lipid bilayer.

1.1.1. Biological and pharmacological importance

Transmembrane proteins are ubiquitous and make up ~30% of the human proteome (1). They are involved in a variety of vital biological roles in cells, for example the relay of signals (2) or transport of ions (3) and larger solutes across the membrane (4), the catalysis of chemical reactions at the membrane (5) or the provision of nodes/junctions for intra- and extracellular scaffolding or cell contacts (6). It is therefore not surprising that integral membrane proteins are of particular interest for the pharmaceutical research and industry (7,8). This is illustrated by the fact that up to 60% of all approved therapeutic drugs act on this class of proteins (7). The dominance of integral membrane proteins as drug targets is not only a consequence of their biological importance but also of their localisation at the membrane surface. This is a hallmark of druggability as it makes them easily accessible to drugs (9) . Within membrane proteins, G-protein coupled receptors (GPCRs) have a privileged position as drug targets since ~35% of all approved small-molecule drugs act on GPCRs or GPCR-related proteins alone (8).

1.1.2. Challenges of working with membrane proteins

Despite the biological and pharmacological importance of integral membrane proteins, their biochemical and biophysical characterisation lacks far behind that of soluble proteins (10). For example, to date only ~3% of all structural data deposited in the protein data bank (PDB) describes membrane proteins (11). Why is that? The very nature that makes many integral membrane proteins key players of cell biology and prime drug targets, their embedment in the lipid bilayer, hampers the research in this field. Membrane proteins are intrinsically hydrophobic and need to be extracted out of the lipid

bilayer for functional and structural studies that typically take place in aqueous solution (Figure 1.1).

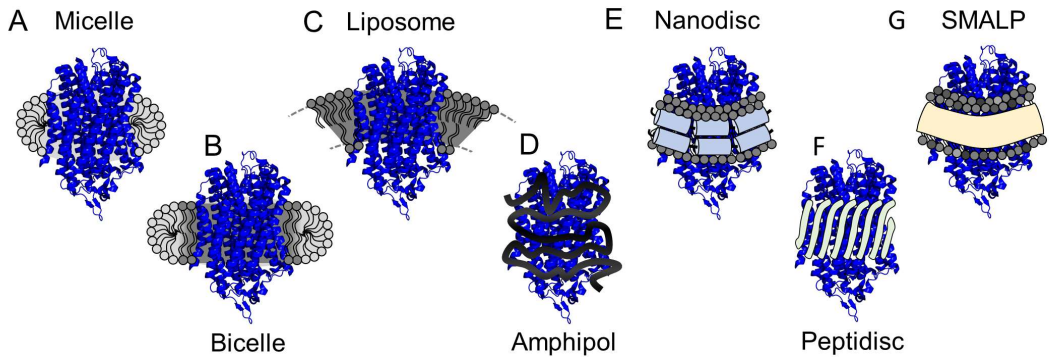


Figure 1.1: Schematic representation of protein solubilisation methods. (A) Protein solubilised in a detergent micellce. (B) Protein solubilised in a bicelle. (C) Protein solubilised in a unilamellar liposome. (D) Protein solubilised by an amphipol. (E) Protein solubilised in a nanodisc. (F) Protein solubilised in a peptidisc. (G) Protein solubilised by styrene-maleic acid (SMA) co-polymers. Detergent molecules are shown in light grey, lipid molecules are shown in dark grey and integral membrane proteins are coloured blue.

Detergent solubilisation is arguably the most common and successful method for membrane protein extraction, but it has some limitations (Figure 1.1A) (12). Detergent micelles are only poor mimics of lipid bilayers as their physicochemical properties differ substantially from the natural lipid environment (12). For example, detergent micelles are highly curved spherical or ellipsoidal objects, whereas lipid bilayer are relatively flat structures (13). Intrinsic curvature is an important modulator of protein folding kinetics, stability and function alongside various other physical descriptors such as the hydrophobic mismatch or bilayer thickness and fluidity (14–18). Furthermore, lateral pressure of lipid bilayers is important for protein function (18–21). This is thought to influence the conformational state of integral membrane proteins that undergo a depth-dependent variation in the cross-sectional protein-lipid interface (18–21). Similarly, lateral tension also affects protein function; for example by controlling the equilibrium of lipid molecules exchanging between the lipid bilayer and allosteric sites (lipid nano-pockets) in mechanosensitive ion channels (22,23).

To date, most “low hanging fruits” of membrane protein research have already been picked so alternative extraction approaches that better mimic the lipid bilayer such as bicelles (24), liposomes (25), amphipols (26), peptidiscs (27) and nanodiscs (28) are more frequently used to investigate more challenging targets (Figure 1.1). A detailed review of the different solubilisation techniques is beyond the scope of this thesis and can be found elsewhere (29). In brief, bicelles are lipid bilayers that are edge-stabilised by detergent molecules (Figure 1.1B). Liposomes are spherical lipid bilayer vesicles that (ideally) do not contain detergent and are mainly used in functional studies (Figure 1.1C). Amphipols are, as the name implies, amphiphilic polymers that wrap around hydrophobic

protein regions to stabilise membrane proteins in aqueous solution (Figure 1.1D). Nanodiscs are synthetic membrane mimics that use amphipathic membrane scaffolding proteins to hold a disc-shaped lipid bilayer together into which membrane proteins incorporate (Figure 1.1E). Peptidiscs consist of multiple short amphipathic bihelical peptides ("one-size-fits-all") that cover hydropobic transmembrane regions to allow solubilisation (Figure 1.1F). All of these extraction strategies have in common that they typically reply on initial detergent-solubilisation in which specific protein-lipid interactions are perturbed due to the removal of native lipids during the process (24–28). The lack of specific protein-lipid interactions may render the protein unstable or inactive as demonstrated in various studies (30–34), even in more "membrane-like" environments.

In 2009, a novel extraction platform based on styrene-maleic acid lipid particles (SMALPs) was developed (Figure 1.1 G) to avoid detergent-solubilisation, and the issues that come with it, entirely (35). This method allows the extraction of integral membrane proteins together with their local lipid environment, which preserves specific protein-lipid and most non-specific protein-bilayer interactions throughout the process (35). However, the usage of SMALPs for protein extraction is limited by their tendency to form heterologous complexes, their susceptibility to low pH and to divalent cations, which can lead to compatibility issues with downstream experimental requirements (35,36). These issues are addressed in combinational approaches that use SMALPs for protein extraction and replace them with more appropriate systems for later applications (37). Nevertheless, some proteins remain intrinsically unstable when extracted out of their natural environment independent of the method chosen.

This raises the question about what defines stable proteins, which has been investigated extensively as the answer to it might allow scientists to engineer robust protein variants that behave better in non-physiological environments. Even after 50 years of research in this field, we still struggle to accurately predict protein stability due to its complexity that is governed by the variety and combinations of parameters known to be involved (38). Some of the best discriminators of thermostable proteins are a more hydrophobic core-region, a higher number of salt-bridges and more compact protein packing, for example *via* shortened loops (38). However, the energy-scoring functions of prediction tools, which represent "rules" for protein stabilisation, were mostly derived from data of soluble proteins and typically neglect the effects of lipids or the lipid-bilayer on protein stability (39). Therefore, scientist still rely on time-consuming and resource-intensive trial-and-error approaches to engineer stable integral membrane proteins (40). The arguably most common and successful method for protein stabilisation is scanning mutagenesis in which the sequence space is systematically probed by alanine, leucine or cysteine substitutions (41). Alanine is often the amino acid of choice due to its

compactness and high helix propensity (42). Other approaches, such as directed evolution by random mutagenesis using error prone polymerase chain reaction (PCR) have also been successfully employed in stabilisation campaigns in the past. However, they suffer from biased changes (adenine for thymine base substitutions are more common than cytosine for guanine substitution) and the requirement for robust and high-throughput assays to select successful mutations (43,44).

A more simple and potentially as effective approach for integral membrane protein stabilisation is the addition of lipids after detergent solubilisation to reinstate physiological protein-lipid interactions (45,46). This strategy has been key for the structural characterisation of a broad range of targets such as ion pumps (47), ion channels (48), electron transport complexes (49), transporters (50) and GPCRs (51). It also requires screening to identify beneficial lipids in many cases, but this process is less laborious than mutagenesis. A much bigger issue is the financial burden that comes with screening a large lipid space as purified lipid is typically very expensive.

1.1.3. X-ray crystallographic structural studies of membrane proteins

Structural biology is mainly based on three techniques: X-ray crystallography, cryogenic electron microscopy (cryo-EM) and nuclear magnetic resonance (NMR). X-ray crystallography has been historically most successful and constitutes ~80% off all protein structures deposited in the PDB (https://www.rcsb.org/stats/summary). Cryo-EM recently exploded in popularity due to technical advances in the field enabling routine single-molecule structure determination, i.e. no requirement for crystals, at near-atomic resolution (52). NMR can also probe samples in solution for structure determination but only plays a minor role in this field due to the upper target size limitations and long data collection times (53). In the following, I focus on the structure determination of membrane proteins using X-ray crystallography, which is most relevant to this project, and give a brief overview of the methodology, challenges and advances of this technique in the context of membrane proteins. A more detailed description of the methodology is beyond the scope of this thesis and can be found elsewhere (54).

X-ray crystallography is used to determine the three-dimensional (3D) structure of molecules by recording the diffraction of an incident X-ray beam into different directions after hitting the sample (54). This technique requires the sample to be crystalline as photon scattering from electrons of single molecules gives a too weak signal to noise ratio and data processing is further complicated by random sample orientations (54). In crystals, individual molecules are arranged in a regularly ordered lattice in 3D space (54). When X-rays hit a crystalline sample, constructive interference is recorded as diffraction spots (reflections) on a detector if the waves remain in phase, that is, if the travel path between different scattering events (one per photon) is equal to integer multiples of the

wavelength (Figure 1.2) (54). Therefore, a well-ordered crystal acts as a signal amplifier. This is mathematically described by Bragg's law, which relates the wavelength, angle of the incident beam and the crystal lattice spacing to constructive interference of X-ray scattering (55). The position and pattern of reflections on a detector yields information about the dimension of the simplest repeating unit constituting the crystal by translation in all three space directions (unit cell) and the internal arrangement of scattering motifs within the unit cell (symmetry) (54). The intensity of each recorded reflection contains structural information about every single scattering motif in the unit cell. It is determined by the relative positions of atoms to the unit cell origin, the fraction of every equivalent position (occupancy) and the efficiency of X-ray scattering by each atom or, more specifically, by the group of electrons distributed around the atom nucleus (54). This information can be used to reconstruct an electron density map of the unit cell by Fourier transformation into which a model is then built to obtain the structure of the crystallised target (Figure 1.2).

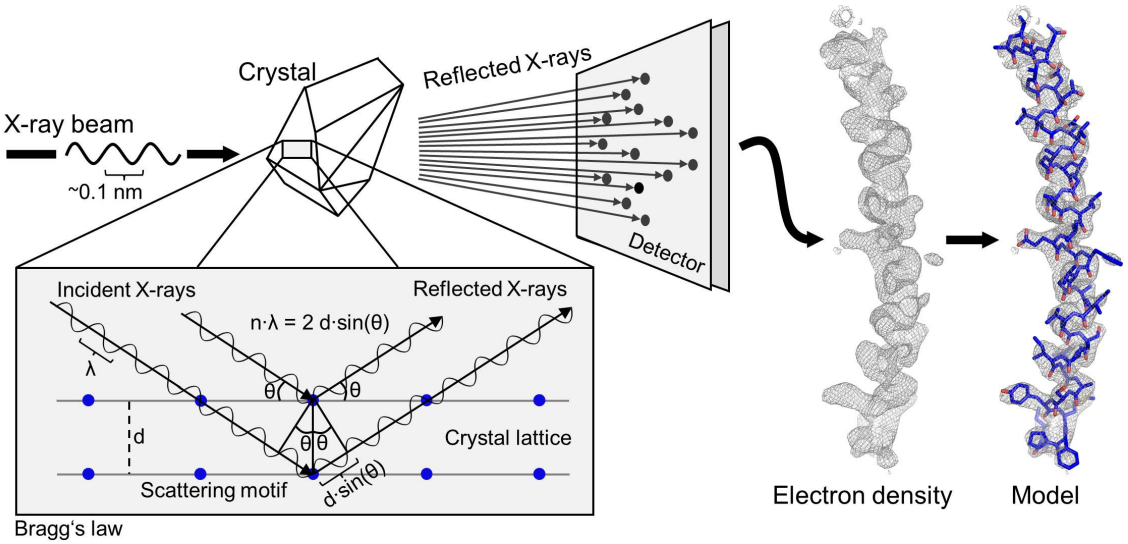


Figure 1.2: Schematic represnetation of an X-ray diffraction experiment. The X-ray beam is diffracted by the electrons of scattering motifes (protein molecules) in a crystal according to Bragg's law (λ=wavelength, n=integer number, d=crystal lattice spacing, θ=angle of incident/reflected light) . The diffraction pattern is recorded on a detector and then used to reconstruct an electron density map of the crystal contents for model building.

The last paragraph introduced the importance of crystals for X-ray crystallography. Therefore, it is also important to note that the solubilisation and crystallisation method used directly affects crystal formation. For example, the use of detergents generally yields so called type-II crystals in which crystal contacts form at polar protein regions that protrude from the micelle (Figure 1.3) (56). This arrangement leads to rather large solvent channels in the crystal, *i.e.* a high solvent content, to a loose crystal packing, and often, to poor diffraction (56). Short chain detergents can promote the formation of crystal contacts at polar regions and support tighter packing, but are at the same time harsh on the protein (57). Alternatively, fusion-proteins (*e.g.* b_{562}RIL) or selective

antibodies/nanobodies are often used to replace flexible protein regions (58) or increase the hydrophilic surface area, and thus the likelihood of crystal contact formation (59). Antibodies and nanobodies have the added benefit of providing conformational stabilisation, which further helps crystal formation (59). Type-I crystals are dominated by hydrophobic interactions that drive the formation of planar membrane-like arrays. These stack on top of each other, which ultimately leads to a tighter packing as compared to type-II crystals (Figure 1.3) (56). This has three distinct advantages. The first one is that tightly packed crystals are more stable, which simplifies crystal handing. The second and third advantage is that tightly packed crystals have lower solvent contents and are typically better ordered, which both improves diffraction quality (60). In general, crystallisation following *in surfo* methods (methods using protein-detergent complexes) yields type-II crystals, whereas *in meso* methods (method using a lipidic mesophase) produce type-I crystals (Figure 1.3).

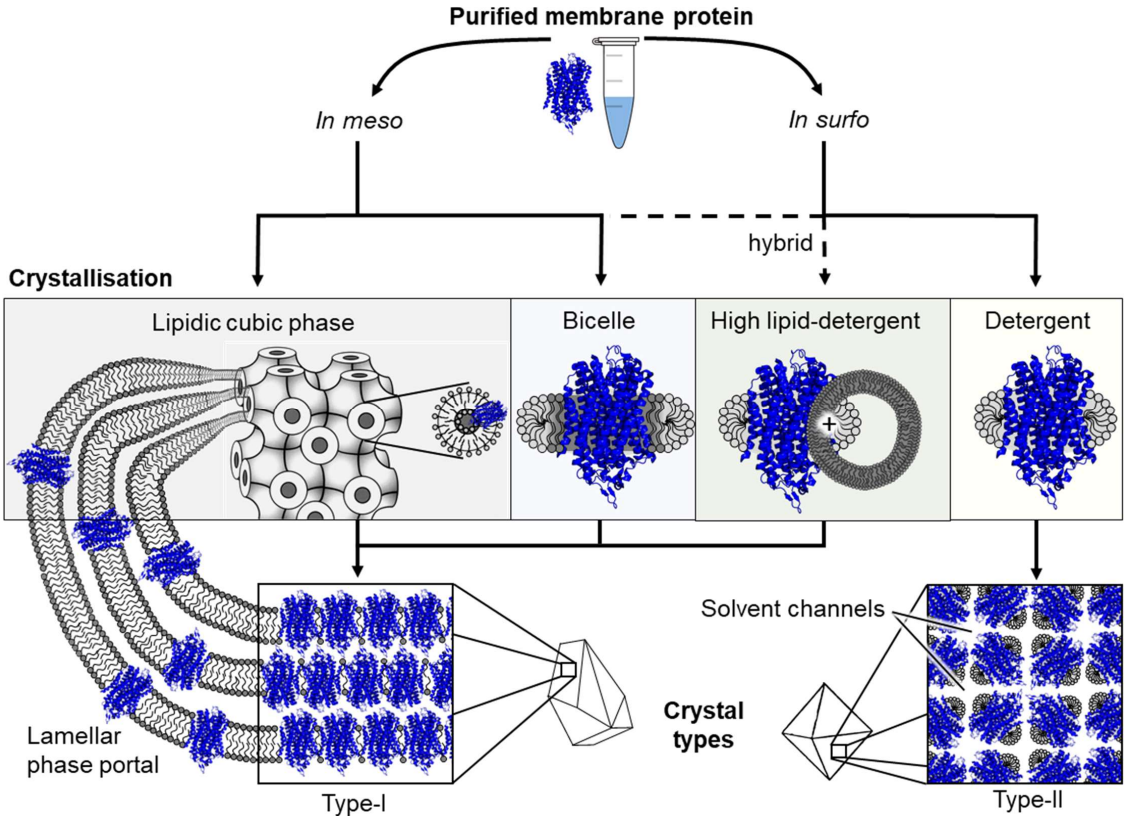


Figure 1.3: Schematic overview of *in meso* and *in surfo* crystallisation methods leading to type-I and type-II crystal formation, respectively. Lipidic cubic phase representation adapted from Landau and Rosenbusch (61). Detergent molecules shown in light grey, lipid molecules in dark grey and integral membrane proteins in blue.

In recent years, the lipidic cubic phase (LCP) crystallisation from bicontinuous mesophases has become the gold standard for high-resolution structural studies of membrane proteins by X-ray crystallography (61). It proved key to the structure solution of many challenging targets, for example GPCRs (62). As *in meso* method, LCP crystallisation yields favourable type-I crystals (Figure 1.3). LCP is a gel-like liquid

crystalline phase that requires a defined water to lipid ratio and temperature for its spontaneous formation (63). It is characterised by a single, highly curved lipid bilayer separating space into two continues (bicontinuous) non-intersecting networks of solvent channels (63). Cis-monosaturated monoacylglycerols (MAGs) are most commonly used as host lipids in LCP as they can be easily modified to influence the phase behaviour/properties according to project needs and form all physiologically relevant water- or temperature-induced mesophases (64). 9.9MAG is the most prominent host lipid used for structural studies in LCP (Figure 1.4), but the choice of lipid is protein specific and often requires screening. For example, shorter-chain MAGs such as 7.7MAG are particularly suited for targets with large extramembrane domains as they form bigger solvent challenges in LCP, which promotes the free diffusion of protein throughout the bicontinuous mesophase (65). Free protein movement in LCP is key to crystallisation since crystal growth is thought to rely on a continuous supply of protein molecules from bulk LCP to a precipitant-induced multilamellar phase portal (Figure 1.3), where nucleation takes place (63).

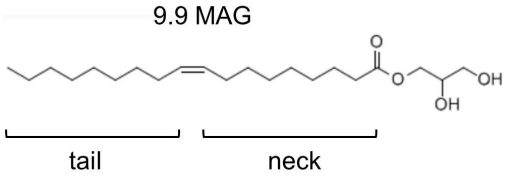


Figure 1.4: N,T-MAG nomenclature. N denotes the number of carbon atoms between the polar head group, also called neck, and the double-bond, whereas T described the tail lengths as number of carbon atoms following that double bond.

LCP not only promotes the formation of type-I crystals, but is also a better mimic of the natural lipid-bilayer than detergent micelles, which is reflected in a greater protein stability (66). Moreover, the natural tendency of membrane proteins to reconstitute from detergent micelles into the more native-like LCP upon mixing can be used for protein concentration in a stabilising environment (67). This is called Cubicon method and practically done in several rounds of spiking LCP with additional protein sample, leading to the formation of a turbid dispersion of hydrated LCP into which protein is reconstituted from detergent micelles. Afterwards, the protein-depleted aqueous phase is removed and some more lipid added to return to optically clear LCP (67).

Despite all these advantages LCP did not replace *in surfo* methods as it is technically challenging and requires specialised tools and robotics due to its gel-like viscosity. High lipid-detergent crystallisation (HiLiDe) combines the advantages of working in a more native-like lipidic environment with the simplicity of *in surfo* crystallisation (68). In this hybrid method, protein is mixed with high concentrations of lipid and detergent prior to setting up routine vapour-diffusion crystallisation trials. In contrast to LCP, virtually any lipid can be used for screening and the sample remains liquid, which greatly simplifies handling (68). At the same time, HiLiDe crystallisation also yields favourable type-I

crystals (Figure 1.3) (68). However, the lipid-detergent screening requires large amounts of sample, which is a limiting factor when working with membrane proteins (68).

Taken together, the last paragraph clearly illustrates that there is no “silver bullet” for membrane protein crystallisation. Instead, an individual target-based choice must be made and often more than one method is used to maximise chances of getting good diffracting membrane protein crystals.

1.2. Introduction to pyrophosphatases

Pyrophosphatases (PPases) catalyse the hydrolysis of inorganic pyrophosphate (PP_i), a by-product of nearly 200 biosynthetic reactions, across all kingdoms of life (69,70). They were first discovered in 1928 in several mammalian tissues (71,72). The removal of excess PP_i is crucial to drive vital biological processes in which it forms, for example DNA/RNA polymerisation, tRNA charging and polysaccharide synthesis (69,70). Moreover, PP_i has a regulatory role for a variety of enzymes without participating in the reaction itself (69,70). Consequently, PPase activity broadly affects cell metabolism.

1.2.1. Soluble pyrophosphatases

The intracellular PP_i recycling is dominated by soluble PPases (S-PPases), which are subdivided into two major families (family I and II) that are evolutionary unrelated (73). They differ in their sequence, structure and catalytic mechanism (73). Membrane-bound PPases (M-PPases), another class of PPases, play only a minor role for PP_i recycling but have broader biological functions as primary ion-pumps (1.2.2). They are also evolutionary unrelated to family I and family II S-PPases (73).

In 1978, the first ever S-PPase structure was solved from *Saccharomyces cerevisiae* (*S. cerevisiae*) (74). It belongs to family I, which are ubiquitous, single-domain enzymes with a 5-strand β-barrel that adapts an OB-fold in the protein core region (Figure 1.5 A) (74). Family I S-PPases typically organise as dimers in eukaryotes (75) and hexamers in prokaryotes (76). The active site sits on top of the core region and is comprised of a conserved D-(S/G/N)-D-P-X-D-X-X motif (X=C/I/L/M/V) (75). In the substrate-bound conformation three to four metal ions are present at the active site of which one to two are brought in by PP_i and two are held in place by D120 (*S. cerevisiae* numbering) of the conserved motif (77,78). A second aspartate of the conserved motif, D117, is key to activate water for the nucleophilic attack on PP_i together with coordinating metal ions M1 and M2 (Figure 1.5 B) (77,78).

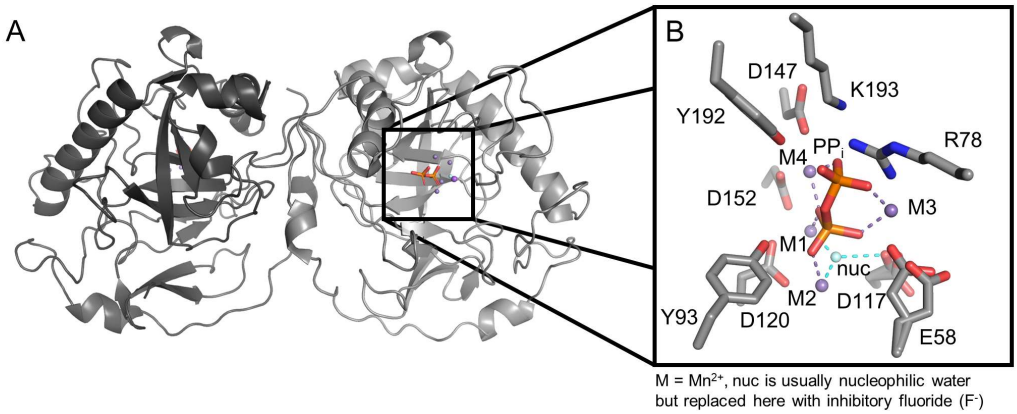


Figure 1.5: Homodimeric type-I S-PPase from *S. cerevisiae*. Subunits coloured in different shades of grey. (A) Mn_4PP_i-bound structure with inhibitory fluoride (cyan sphere) at the active site. (B) Black box shows magnification of the active site with metal ions (here Mn^{2+} as magenta sphere) and PP_i coordinating residues shown. Metal cage and nucleophile coordination is indicated by magenta and cyan dashed lines, respectively

The first structures of family II S-PPases were solved in 2001 from *Streptococcus mutans* (*S. mutans*), *Streptococcus gordonii (S. gordonii)* and *Bacillus subtilis* (*B. subtilis*) (79,80). They organise as dimers and are less ubiquitous than family I S-PPases (mainly found in *Clostridia* and *Bacilli*) (73). As mentioned earlier, family II S-PPases are structurally distinct from family I S-PPases and consist of a N-terminal and C-terminal domain, each with a 5-strand β-sheet and several α-helices (Figure 1.6A) (79–81). The active site is formed at the domain interface and contains four conserved aspartate and two conserved histidine residues (79–81). Their D-H-H motif assigns them to the DHH phosphoesterase superfamily (73). The presence of histidine in the active site leads to a preference for transition metal cations such as Mn^{2+} and Co^{2+} over Mg^{2+} (79–81). Of the up to four metal ions present in the substrate-bound form, three (M1, M2, M3) coordinate the nucleophilic water that is poised to attack PP_i (Figure 1.6B) (79–81).

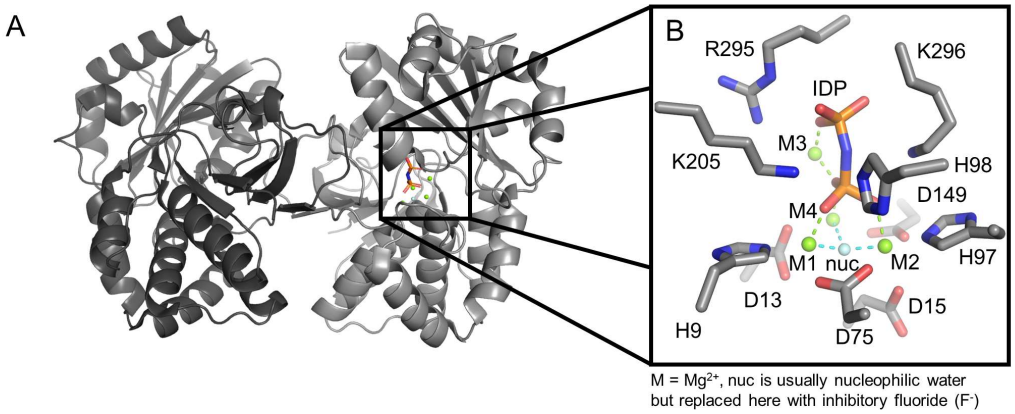


Figure 1.6: Homodimeric type-II S-PPase from *B. subtilis*. Subunits coloured in different shades of grey. (A) Mg_4PP_i-bound structure with fluoride ion (cyan sphere) at the active site. (B) Black box shows magnification of the active site with metal ions (here Mg^{2+} as green spheres) or PP_i coordinating residues shown. Metal cage and nucleophile coordination is indicated by green and cyan dashed lines, respectively.

The structural differences at the active site between family I and family II S-PPases give rise to different reaction mechanisms and can explain why the catalytic turnover (k_{cat}) by family II S-PPases is about 10-fold faster (k_{cat}: ~2000 s^{-1}) compared to family I S-PPases (k_{cat}: 200 s^{-1}) (73). Both reactions take place in a metal-phosphate cage without an enzyme-phosphate intermediate upon nucleophile generation and leaving group activation (81–83). However, the chronological order of nucleophilic attack and leaving group departure is thought to differ between the two families. In family II S-PPases, the tri-metal coordination of nucleophilic water drives its deprotonation more efficiently compared to the coordination by only two metal ions in family-I S-PPases and gives reason for their improved reaction kinetics (81). Moreover, substrate is distorted at the active site of family II S-PPases and H98 (*B. subtilis* numbering) is capable of donating a full proton to the bridging oxygen of PP_i for leaving group activation (Figure 1.6B) (81). This implies a dissociative mechanism in which the leaving group departure precedes the nucleophilic attack *via* formation of a hydrated PO_3^- anion (metaphosphate). There is no such proton donor at an equivalent position in family I S-PPases (Figure 1.5 B). Their leaving group is stabilised by protein side chains that coordinate every lone electron pair of PP_i and can only transfer a "partial proton". Therefore, the nucleophilic attack must follow an associative mechanism in which the leaving group departs only after the nucleophilic attack and formation of a pentavalent phosphorus transition state (83). Both, family I and family II S-PPases, are inhibited by fluoride, which replaces the nucleophilic water but is only a weak nucleophile itself due to its urge to abstract protons and retain electrons instead of donating electrons to a new covalent bond (73).

1.2.2. Membrane-bound pyrophosphatases

M-PPases are yet different to family I and family II S-PPases. They do not only differ in their sequence, structure and catalytic mechanism, but also in their function (73). M-PPases were first found 1966 in the photosynthetic bacterium *Rhodospirillum rubrum* (*R. rubrum*) (84). Shortly afterwards, their role in linking the P-O-P anhydride hydrolysis/synthesis to H^+ translocation over the membrane was discovered (85). Thus, M-PPases are functionally similar to ATPases. In the following years, M-PPases have been found in several species of bacteria, archaea and eukarya, but none in multicellular animals (86–89). It became apparent that PP_i hydrolysis is dominant and drives ion pumping, which is not limited to H^+ but also includes Na^+ (90,91). Accordingly, M-PPases are classified based on their ion selectivity and co-factor dependence (Table 1.1).

To date, H^+-pumping (H^+-PPase), Na^+-pumping (Na^+-PPase) and dual-pumping (Na^+,H^+-PPase) M-PPases have been found (85,90,91). Na^+,H^+-PPases are further subdivided into Na^+-regulated and true dual-pumps, of which only the latter retain their ability to translocate H^+ at supraphysiological Na^+ levels (92). While most M-PPases

require potassium (K^+) for maximal catalytic activity, a subclass of K^+-independent enzymes has been identified among the H^+-PPases (93).

Table 1.1: Classification of M-PPases.

Classification		Mechanistic insights ǂ		Other	Example
Ion selectivity	Cation dependence	Semi-conserved glutamate position	Cationic centre occupation		
Na^+	Na^+, K^+	$E^{6.53}$	K^+	Evolutionary ancestor	*Thermotoga maritima* (*Tm*-PPase)
H^+	K^+	$E^{6.57}$	K^+	$E^{6.57} \rightarrow E^{5.43}$ in *Flavobacterium johnsoniae*	*Vigna radiata* (*Vr*-PPase)
	-	$E^{6.53}$	$K^{12.46}$	Coupled $A^{12.46}K$ and $G/A^{12.49}T$ change	*Pyrobaculum aerophilum* (*Pa*-PPase)
Na^+,H^+	Na^+, K^+	$E^{6.53}$	K^+	Na^+-regulated	*Bacteroides vulgatus* (*Bv*-PPase)
	Na^+, K^+	$E^{6.53}$	K^+	True dual-pump	*Clostridium leptum* (*Cl*-PPase)

ǂ Ballesteros & Weinstein nomenclature where the first number denotes helix and second number offset of well conserved residue in the centre of its helix(at position 50 as per definition) (94)

Na^+-pumping M-PPases including dual-pumps are exclusive to prokaryotes and predominately localised in the cell membrane, whereas H^+-PPases are also found in eukaryotes were they occur most frequently in organelle membranes (92,95). In general, the direction of ion transport is from the cytosol, where the PP_i hydrolysis takes place, to the periplasmic space/exterior, if localised in cell membranes, or the lumen if embedded in organelle membranes.

1.2.3. Biological role of M-PPases in plants, bacteria and parasites

M-PPases are important for cell survival under stress conditions such as low-light, nutrient and oxygen deficiency, drought, cold or hyper salinity (96). This is because they recycle part of the energy stored in P-O-P anhydride bond of PP_i to establish a membrane potential and energise secondary transporters (96). Consequently, M-PPases are common in bacteria that populate harsh environments such as the obligate anaerobe *Bacteroides* and deep-sea organisms (96). In line with this, their overexpression was shown to be suitable for engineering more robust bacterial strains. For example, the heterologous introduction of Na^+,H^+-PPases in *Escherichia coli* (*E. coli*) plasma membranes yields more salt tolerant strains, probably due to increased Na^+ export capabilities (97).

In plants, H^+-PPases are abundant in the tonoplast and energise, together with H^+-ATPases, the transport of molecules in and out of the vacuole (98). M-PPase activity is also important for plant hormone homeostasis, and thus, plant maturation. This is because H^+-PPase expression levels directly influence auxin trafficking by carrier proteins, which are mostly ATP or H^+ driven (Figure 1.7) (99). As a result, the knockout of H^+-PPase in *Arabidopsis thaliana* (*A. thaliana*) manifests in a phenotype with impaired root and shoot development (99-103). H^+-PPase overexpression improved root and

shoot development under low-energy stress conditions instead, which enhances water adsorption and therefore drought resistance (104–112). Similarly, high-salt tolerance was conferred by H^+-PPase and Na^+,H^+-PPase overexpression in various plants, likely because the enhanced membrane potential and Na^+-pumping activity promotes solute uptake into the vacuole and thereby prevents water efflux (104–112). Cold resistance is likewise conferred by the transport and accumulation of cryoprotectants in the vacuole (113). The enhanced solute uptake, better water retention and more stable ion homeostasis may also account for reported effects of increased chlorophyll content, photosynthesis, fibre yield and cell membrane integrity in transgenic plants (106–109).

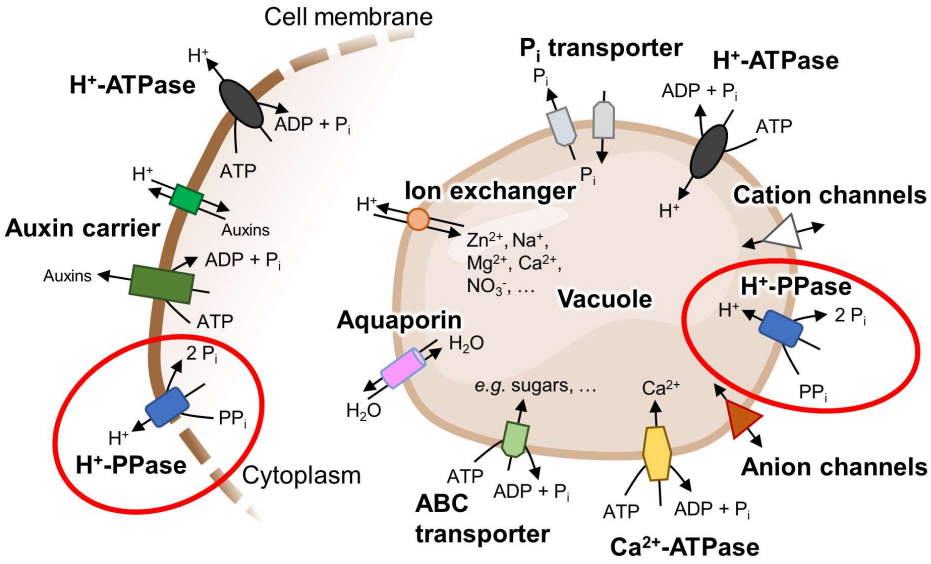


Figure 1.7: Schematic overview of membrane trafficking in plant organelles. Major transport proteins for solute translocation across the tonoplast membrane (in and out of vacuoles) and auxin trafficking at the cell membrane of plants. Figure shows only the key players of membrane trafficking and organelles do not necessarily contain all of the depicted proteins throughout all species. H^+-PPases are highlighted by a red circle.

In parasites, H^+-PPases mostly occur in the membrane of acidocalcisomes (Figure 1.8), which are acidic storage compartments for cations and phosphorus compounds such as orthophosphate (P_i), poly phosphate (poly-P) and PP_i (114,115). Phosphate and cation removal from, or release into, the cytoplasm is crucial for the osmotic homeostasis, which is of particular importance for parasites due to their transitioning between vector and host system environments (116,117). H^+-PPases drive, together with H^+-ATPases, the acidification of acidocalcisomes (118). In H^+-PPase knockout strains of *Trypanosoma bruceii* (*T. bruceii*), acidification by H^+-ATPases alone was insufficient to maintain functional integrity (118). Poly-P levels dropped 10-fold and Ca^{2+} transport, mediated by the Ca^{2+} chelating potential of poly-P and H^+/Ca^{2+}-exchanging ATPases, was impaired (118). Both are key to adapt to osmotic stress. For example, in hypoosmotic environments the hydrolysis of poly-P in, and release of Ca^{2+} from, acidocalcisomes is triggered to maintain the internal ionic strength (116,117). The acidification of

acidocalcisomes also mediates the cellular pH regulation, likely by removal of H^+ from the cytosol under acidic conditions or release from the acidocalcisome under alkaline conditions (118). Consequently, H^+-PPase knockout strains of *T. bruceii* also failed to maintain their physiological pH following extracellular pH changes (118).

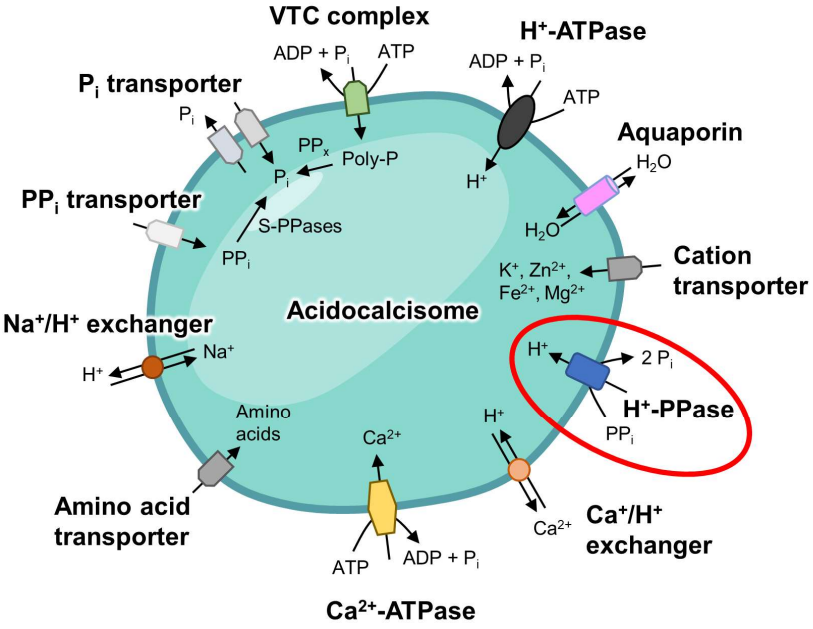


Figure 1.8: Schematic overview of membrane trafficking in parasite organelles. Major transport proteins involved in the solute import/export into and from the acidocalcisomes of parasites. Figure shows only the key players of membrane trafficking and organelles do not necessarily contain all of the depicted proteins throughout all species. H^+-PPases are highlighted by a red circle.

1.3. Structural features of M-PPases

1.3.1. Structural overview

M-PPases are large (66-89 kDa), single-domain integral membrane proteins comprising of two identical monomers, each with 15-17 transmembrane helices (Figure 1.9A). Prior to the availability of structural data, the mechanistic investigation of M-PPase biochemistry relied mostly on functional studies of variant protein, which on their own only provided limited in-depth insights (119–126). This changed in 2012 when structures of the K^+-dependent H^+-PPase from *Vigna radiata* (*Vr*-PPase) and the K^+-dependent Na^+-PPase from *Thermotoga maritima* (*Tm*-PPase) were solved, enabling a detailed structure-function analysis (127,128). We significantly advanced our mechanistic understanding of M-PPase biochemistry since then by mapping different conformational states (129–131), but still miss subclass diversity in our structural data as all efforts solely focused on *Tm*-PPase and *Vr*-PPase (Table 1.2). Nevertheless, the overall structures of M-PPases are likely similar across all subclasses due to their high sequence identity (23-56%) (95). Indeed, structural alignments of *Tm*-PPase and *Vr*-PPase support this (Table S1).

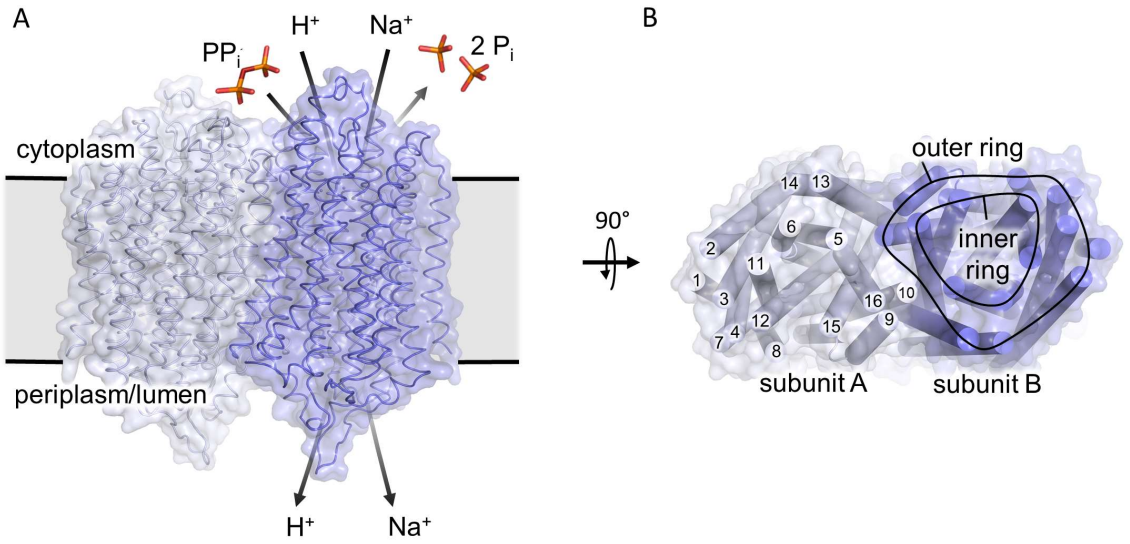


Figure 1.9: Structural overview of M-PPases. (A) Homodimeric M-PPase viewed from the membrane plane and schematic representation of PP_i-energised ion translocation. (B) Concentric ring arrangements of transmembrane helices viewed from the cytoplasmic site. Loops were removed for clarity. Helices are numbered in subunit A and concentric ring arrangements are highlighted in subunit B.

Table 1.2: Overview of M-PPase structures.

Classification	PDB ID	Name	Res.‡ (Å)	Details ∓	State	Ref.‡
K⁺-dependent Na⁺-PPase (T. maritima)	5LZQ	*Tm*-PPase:Mg_5IDP	3.5	AS: 1 IDP, 5 Mg^{2+}, 1 K^+ IG: 1 Na^+	*IDP-bound*	(129)
	4AV6	*Tm*-PPase:Mg_4P_{i2}	4.0	AS: 2 P_i, 4 Mg^{2+} IG: -	product-bound	(129)
	5LZR	*Tm*-PPase:Mg_2WO_4	4.0	AS: 1 WO_4^{2-}, 2 Mg^{2+} IG: -	relaxed product-bound	(129)
	4AV3	*Tm*-PPase:CaMg	2.6	AS: 1 Ca^{2+}, 1 Mg^{2+} IG: -	resting	(127)
	6QXA	*Tm*-PPase:ATC_2	3.7	AS: 1 IDP, 5 Mg^{2+} IG: 1 Na^+ $Loop_{6-7}$: asym. 2 ATC	asym. locked	(131)
K⁺-dependent H⁺-PPase (*V. radiata*)	4A01	*Vr*-PPase:Mg_5IDP	2.4	AS: 1 IDP, 5 Mg^{2+},1 K^+ IG: structural water	IDP-bound	(128)
	5GPJ	*Vr*-PPase:Mg_2P_i	3.5	AS: 1 P_i, 2 Mg^{2+} IG: -	relaxed product-bound	(129)
	6AFS	*Vr*-PPase:Mg_5P_{i2}	2.3	AS: 2 P_i, 5 Mg^{2+}, 1 K^+ IG: -	product-bound	(130)
	6AFT-6AFZ		2.2-2.8	AS: 2 P_i, 5 Mg^{2+}, 1 K^+ IG: $E^{6.57}Q$ EC: $L^{12.64}M/K$, $T^{5.35}D$, $E^{5.33}A/S/H$	product-bound	(130)

∓ Binding of ligands and ions to the AS (active site), IG (ion gate) and EC (exit channel) of each subunit unless indicated otherwise (asym.: asymmetrical binding to one subunit only).
* Root mean square deviation (rmsd) of structural alignment based on Cα atoms of subunit A.
‡ Res.: Resolution, Ref.: References.

In general, the helices of each subunit arrange into an inner ring (helices 5-6, 11-12, 15-16) containing the functional core and an outer ring (helices 1-4, 7-10, 13-14) of largely unknown function (Figure 1.9B). In the following, I will introduce the functional core regions in more detail and point out unanswered questions in the field, of which some were addressed in this PhD project (1.7). The residue numbering scheme I follow throughout the thesis ($X^{Y.Z}$) is adopted from the Ballesteros & Weinstein nomenclature that is used for easy comparison of residue conservation/identity in GPCRs (92).

X represents the amino acid as single-letter code, Y denotes the helix on which it is located and Z defines the offset of a well conserved residue (at position 50 as per definition) in the centre of this helix.

1.3.2. Active site

The active site in M-PPases protrudes about 20 Å out of the membrane plane into the cytoplasm (Figure 1.10). It is lined up by an evolutionary conserved environment of aspartate, asparagine and lysine residues that provide the basis for PP_i binding and hydrolysis (127,128). These include $K^{5.58}$, $D^{5.61}$, $D^{5.65}$, $D^{6.35}$, $D^{6.39}$, $D^{6.43}$, $D^{11.57}$, $D^{12.39}$, $N^{12.43}$, $D^{15.61}$, $K^{15.65}$, $D^{16.32}$, $D^{16.35}$, $K^{16.38}$ and $D^{16.39}$ (Figure 1.10A-B). Amino acid substitutions in this region are not tolerated and typically lead to inactive protein (119–121,128,132). The aspartate and asparagine side chains coordinate up to five Mg^{2+}, which capture PP_i in a metal cage, whereas lysin side chains directly stabilise PP_i binding at the active site (Figure 1.10B). Of the five Mg^{2+} present in the imidodiphosphate-bound (IDP) structures (*Vr/Tm*-PPase:Mg_5IDP), two are brought in by the enzymatically active substrate (Mg_2PP_i), two bind to activating high-affinity sites (K_d: ~20-460 µM) and one binds to a putatively inhibitory low-affinity site (K_d: ~100 mM) (133,134). Despite grouping in different M-PPase subclasses (Table 1.2), the IDP-bound active site organisation of *Tm*/*Vr*-PPase are nearly identical with a root mean square deviations (rmsd) of only ~0.37 Å when Cα atoms of each monomer are superimposed.

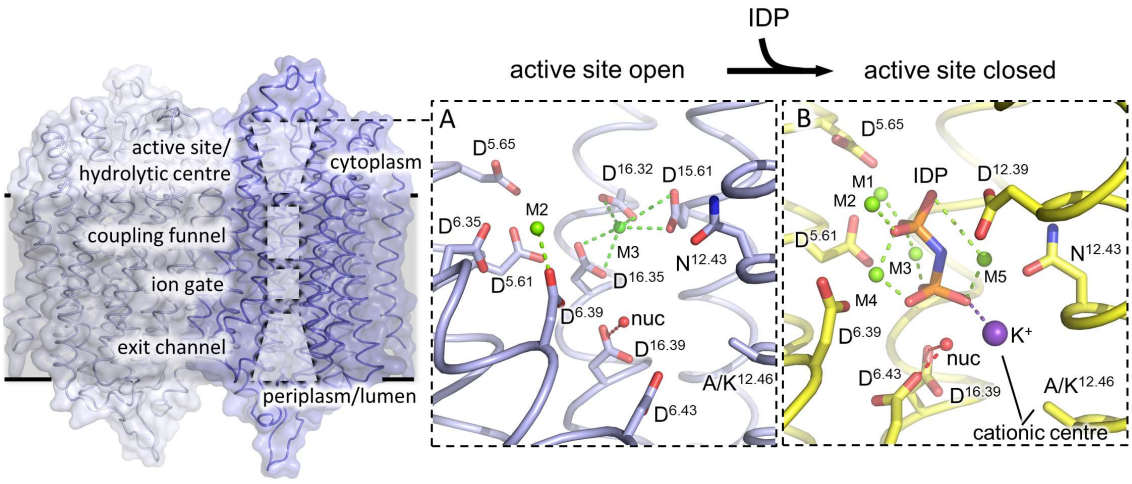


Figure 1.10: M-PPase functional core regions with a focus on the active site. (A) Close-up view of the resting-state active site of *Tm*-PPase with helix 11 removed for clarity. M2 is Mg^{2+}, M3 is inhibitory Ca^{2+} (green spheres). Nucleophilic water of *Vr*-PPase:Mg_5IDP was modelled into the active site. Coordination of metal ions and nucelophilic water shown by dashed lines in green and red, respectively. (B) Close-up view of the active site in the IDP-bound state (*Vr*-PPase:Mg_5IDP) with helix 11 removed for clarity. M1-M5 are Mg^{2+} ions (green spheres) and form a metal cage. K^+ (purple sphere) is part of the cationic centre in K^+-dependent M-PPases (with $A^{12.46}$). Metal ion interactions with IDP are indicated by dashed lines in green/purple. Nucleophilic water (nuc, red sphere) is activated by $D^{6.43}$ and $D^{16.39}$ as illustrated by dashed lines in red. Protein colouring in this thesis follows previous publications that used shades of yellow/ orange for *Vr*-PPase and shades of blue for *Tm*-PPase structures.

IDP is a substrate analogue that has the bridging oxygen atom of PP_i replaced by nitrogen, making it resistant to hydrolysis, and thus, a PPase inhibitor. It is often used in structural studies to trap conformational states for crystallisation. Other non-hydrolysable substrate analogues commonly used in PPase research are bisphosphonate derivates in which the bridging oxygen atom is replaced by a carbon instead (Figure 1.11). Some examples are methylene-diphosphonate (MEDP), amino-methylene-diphosphonate (AMDP) or 1-hydroxyethylidene-1,1-diphosphonate (HEDP), also called etidronate. Ca^{2+} is also a potent inhibitor of PPases as it replaces divalent metal ions required for PP_i binding to the active site , such as Mg^{2+} or Mn^{2+} (127,135).

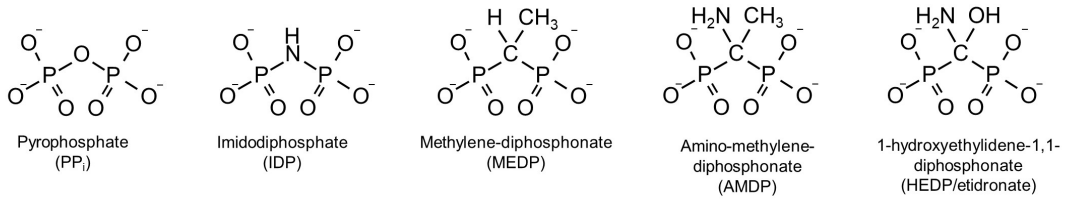


Figure 1.11: Overview of the chemical structure of pyrophosphate and pyrophosphate analogues.

Analysis of the active site configuration in IDP-bound M-PPase structures (Figure 1.10B) and comparison to S-PPases shows that hydrolysis must follow an associative mechanism (129). This is because there is no hydrogen donor side chain nearby the bridging oxygen atom of PP_i, which would allow a metaphosphate formation that could then "pull in" a hydroxide anion for the nucleophilic attack as in family II S-PPases (129). Instead, the nucleophilic water molecule is poised to attack PP_i for reaction initiation and interacts with either one ($D^{6.43}$, resting state) or two aspartates ($D^{6.43}$ and $D^{16.39}$, active state), depending on its activation status (Figure 1.12). This is somewhat similar to the nucleophilic water activation by D117 in family I S-PPases (1.2.1).

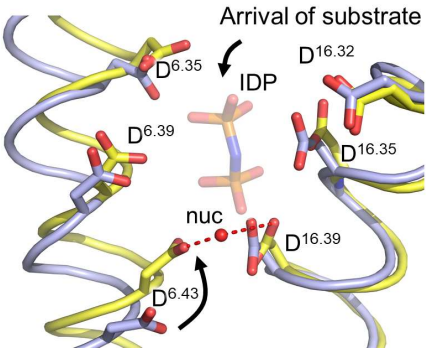


Figure 1.12: Nucleophile coordination in the substrate-bound and resting-state active site of M-PPases. The resting state *Tm*-PPase:CaMg structure is shown blue and substrate-bound *Vr*-PPase:Mg_5IDP is shown in yellow. All helices but helix 6 and 16 were removed for clarity. Arrival of IDP and subsequent reorientation of $D^{6.43}$ is indicated by black arrows. Nucleophilic water is shown as a red sphere and its coordination is indicated by dashed lines.

In contrast to S-PPases, there are no metal ions involved in nucleophile coordination in M-PPases (Figure 1.10B). This explains why M-PPase are not sensitive to fluoride inhibition. Their nucleophilic water has a less anionic character, in other words, $D^{6.43}$ and $D^{16.39}$ are incapable of binding an anion (73). The organisation of PP_i binding and nucleophile activation in M-PPases is also consistent with the fact that they display the slowest substrate turnover (k_{cat}: ~20 s^{-1}) rates of all PPases (73). The polarisation of

water and stabilisation of an hydroxyl anion by metal coordination (lowering the $pK_{a,\ water}$) provides better nucleophile activation than the polarisation by aspartate coordination (73). Additionally, the leaving group activation by the partial or full donation of a proton from protein side chains in S-PPases is more efficient than the encompassing coordination of PP_i by metal ions in M-PPases (77).

The cationic centre is part of the active site and located just below the substrate (Figure 1.10B). It is occupied by K^+ in K^+-dependent M-PPases and thought to be replaced by the ε-NH_3^+ group of $K^{12.46}$ in K^+-independent M-PPases (136). Interestingly, the cationic centre and active site seem to be linked to other functional core regions within and across M-PPase subunits. K^+ binding increases the affinity for Na^+ (90), which binds ~18 Å away at the ion gate (1.3.4). In turn, mutations at ion gate residues lower the catalytic turnover and binding affinity of Mg_2PP_i at the active site (126). Moreover, binding of Mg_2PP_i to one active site changes its affinity for the other active site and binding to both active sites is inhibitory (136). These observations are reviewed in more detail in the context of K^+-(in)dependence and intra-subunit communication in 1.4.1 and 1.5.2, respectively.

1.3.3. Coupling funnel

The coupling funnel bridges the space between the active site, which resides above the membrane plane and the ion gate in the centre of the membrane spanning protein section (Figure 1.13). It couples, as the name implies, PP_i hydrolysis to the transport of Na^+, H^+ or both across the membrane (127,128). A set of highly conserved charged residues including $R/Q^{5.50}$, $D^{6.43}$, $D^{6.50}$, $D/S^{11.50}$, $K^{12.50}$, $K^{16.38}$ and $D^{16.39}$ line up to form an ion translocation pathway through hydrophobic protein regions (Figure 1.13). Of these residues, $D^{6.43}$ and $D^{16.39}$ sit at the interface of the active site, whereas $D^{6.50}$ connects to the ion gate. Amino acid substitution of charged residues of the coupling funnel either inactivate M-PPases entirely or uncouple PP_i hydrolysis (reduced level) from ion translocation (abolished) even when chemical properties are conserved, for example in $D^{6.50}E$ variants (132). Changes to several uncharged amino acids along the coupling funnel (*e.g.* at $I^{12.54}$ or $N^{16.46}$) have similar effects, which further highlights the importance of exact side chain positioning in this region for ion transport (132).

The crystal structure of IDP-bound *Vr*-PPase was of sufficient quality to reliably map a network of charged residue side chains and water molecules that form a Grotthuss chain along which H^+ can be transported (Figure 1.13A). *Tm*-PPase structures were of insufficient quality to map water molecules; however, residue side chains of the coupling funnel appear to orient as seen in *Vr*-PPase when comparing the same conformational states (Figure 1.13B).

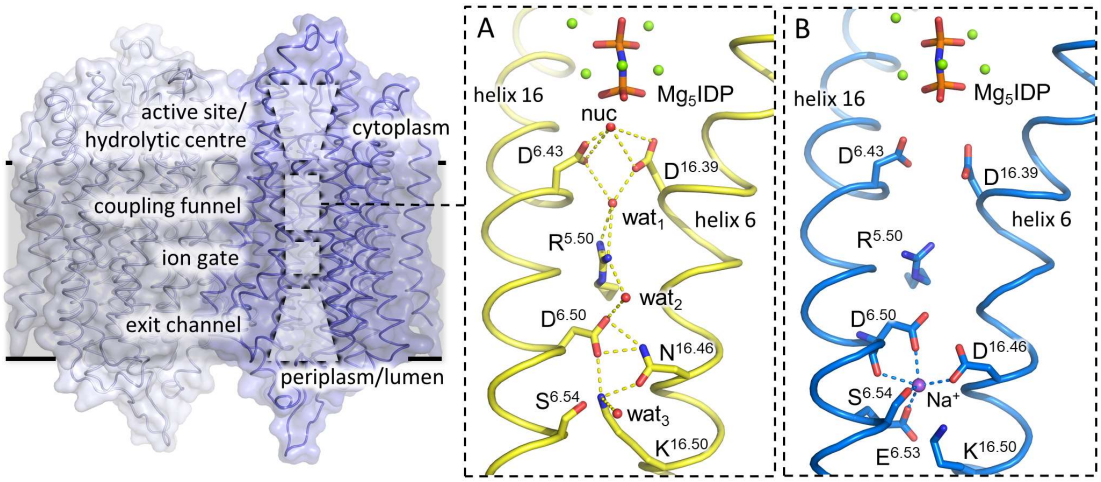


Figure 1.13: M-PPase functional core regions with a focus on the coupling funnel. (A) Close-up view of the coupling funnel of *Vr*-PPase in the IDP-bound conformation with key residues displayed. H^+-translocation pathway including structural waters (nuc+wat_{1-3}, red spheres) shown as dashed lines in yellow. (B) Close-up view of the coupling funnel of *Tm*-PPase in the IDP-bound conformation with key residues displayed. Na^+ coordination at ion-gate shown as dashed lines in blue. The resolution is insufficient to map water molecules.

1.3.4. Ion gate

Ion-pumping selectivity in M-PPases is defined at the ion gate and centred on $K^{16.50}$ (129). In the absence of substrate, $K^{16.50}$ forms an ion triplet ($D^{6.50}$-$K^{16.50}$-$E^{6.53}$) in K^+-dependent Na^+-PPases (Figure 1.14A) or ion pair ($E^{6.57}$-$K^{16.50}$) in K^+-dependent H^+-PPases (Figure 1.14B). Sequence alignments show that the change in ion selectivity is correlated with the downward shift of the glutamate $E^{6.53\rightarrow57}$ by one helix turn in M-PPases (128). Therefore, this shift was initially thought to be the key determinant of ion selectivity and $E^{6.53/57}$ is often referred to as the semi-conserved glutamate in the literature (129). The binding of substrate to the active site leads to the reorientation of helices 6 and 16 and subsequent disruption of the ion-triplet/pair of $K^{16.50}$. This affects K^+-dependent Na^+-PPases with $E^{6.53}$ and K^+-dependent Na^+-PPases with $E^{6.57}$ differently.

In K^+-dependent Na^+-PPases, a Na^+-binding site that was previously occupied by ε-NH^{3+} of $K^{16.50}$ is unmasked. Na^+ is coordinated by $D^{6.50}$, $E^{6.53}$, $S^{6.54}$ and D/$N^{16.46}$ and accounts for the negative charge of $E^{6.53}$ in the centre of the membrane spanning protein section (Figure 1.14A). In K^+-independent H^+-PPases, the reorientation of $K^{16.50}$ does not unmask the Na^+-binding site as it remains coordinated by $D^{6.50}$, $S^{6.54}$ and D/$N^{16.46}$. Instead, a water molecule, which hydrogen bonds to $D^{6.50}$, D/$N^{16.46}$ and $K^{16.50}$, localises nearby. The semi-conserved glutamate, $E^{6.57}$, loses $K^{16.50}$ coordination and remains coordinated by $S^{5.43}$ alone, which cannot account for its negative charge at the ion gate (Figure 1.14B). Therefore, $E^{6.57}$ it is thought to be protonated (129). The nearby water molecule is likely involved in the H^+ transport from the active site through the coupling funnel down to $E^{6.57}$ at the ion gate *via* a hydrogen bond network in a Grotthuss-type mechanism (128).

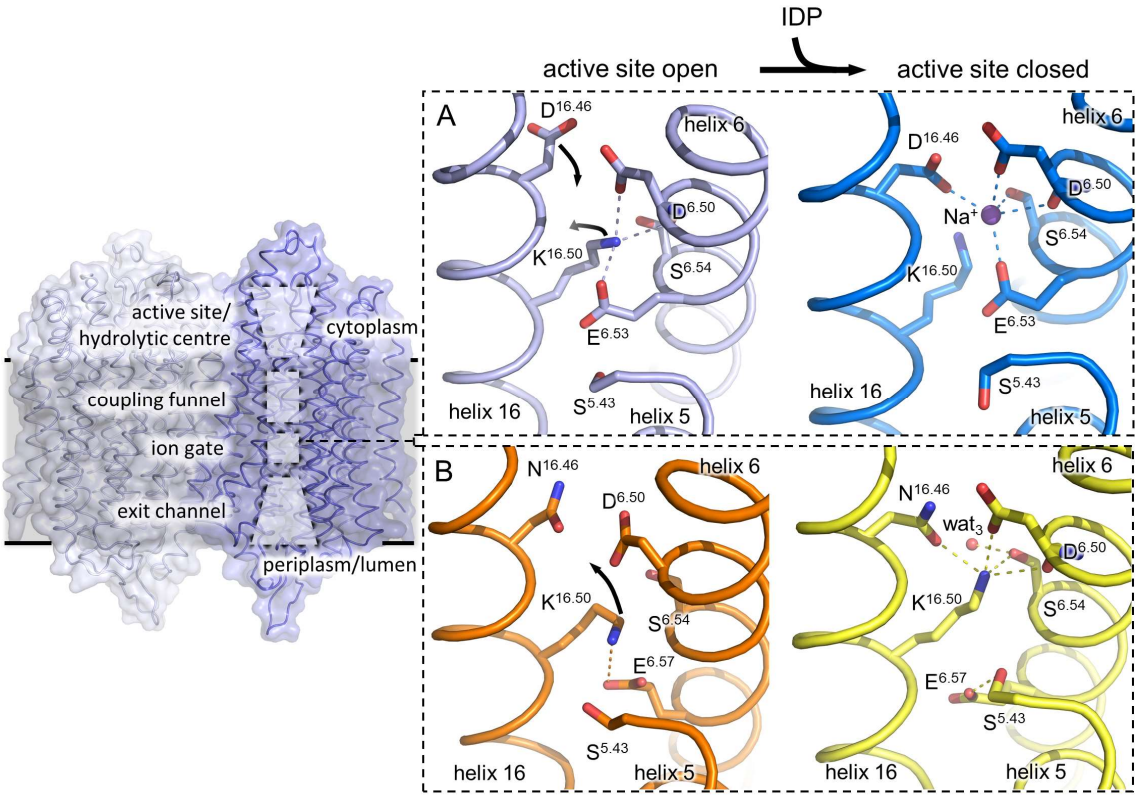


Figure 1.14: M-PPase functional core regions with a focus on the ion gate. (A) Close-up view of the ion gate in K^+-dependent Na^+-PPases. Left panel shows residue configuration when the active site is open (*Tm*-PPase:CaMg). Right panel shows the residue configuration when the active site is fully closed (*Tm*-PPase:Mg_5IDP). Na^+ shown as blue sphere. (B) Close-up view of the ion gate in K^+-dependent H^+-PPases. Left panel shows residue configuration when the active site is open (*Vr*-PPase:MgP_{i2}). Right panel shows the residue configuration when the active site is fully closed (*Vr*-PPase:Mg_5IDP). Dashed lines indicate key interactions and black arrows highlight conformational changes of residue side chains upon substrate binding to the active site. Structural water shown as red sphere.

Taken together, the shift of $E^{6.53}$ (K^+-dependent Na^+-PPases) to $E^{6.57}$ (K^+-dependent H^+-PPases) illustrates a plausible model of ion selectivity in these subclasses as it influences the Na^+-binding site and $E^{6.57}$ protonation state, but what about other M-PPase subclasses? To date, there is not structural data available for K^+-independent H^+-PPases and K^+-dependent Na^+,H^+-PPases that could shed light on the ion selectivity mechanism in these proteins. However, a general model for ion selectivity that focuses on the semi-conserved glutamate position alone is problematic. This is discussed in the context of sequence alignments and M-PPase evolution in 1.4.2.

1.3.5. Exit channel

The exit channel is the least investigated functional core region in M-PPases. It connects to the ion gate on the periplasmic or luminal membrane side and is thought to bind ions less tightly to facilitate their release. This region is very diverse in its amino acid composition and less conserved than the active site, coupling channel or ion gate (130). At the top of the exit channel, just below the ion gate, a set of four mostly non-polar residues ($L^{12.64}$-$X^{5.40}$-$Z^{16.54}$-$Z^{6.61}$; X= G/A/S, Z=L/M/V/F/I/T) form a checkpoint for

ion-release, also referred to as hydrophobic gate (Figure 1.15). Of these, only $L^{12.64}$ is highly conserved and thought to act as a hydrophobic barrier to seal an ion pore and therefore prevent ion backflow (130). Consistent with this idea, mutations at $L^{12.46}$ ($L^{12.46}$A/M/K) primarily affect the coupling of H^+-pumping and PP_i hydrolysis in *Vr*-PPase (130,132).

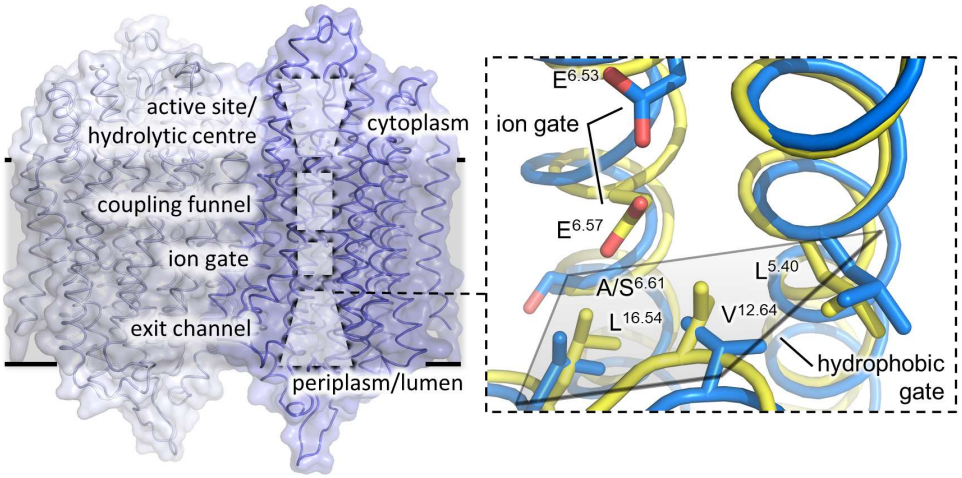


Figure 1.15: M-PPase functional core regions with a focus on the exit channel. Black box shows close-up view of the hydrophobic gate at the top of the exit channel in the IDP-bound *Tm*-PPase (blue)/*Vr*-PPase (yellow) structure. Residues forming the hydrophobic gate are highlighted at the corners of a grey plane, which illustrates a hydrophobic barrier that controls ion flow.

Throughout different catalytic states, the pore size at the hydrophobic gate varies from ~2-4 Å (130). The biggest pore size is observed in the 2 P_i-bound state of *Vr*-PPase and likely mimics an intermediate gate-open state that is thought to form immediately after product formation but could not be stabilised by ligand or inhibitor binding for structural characterisation so far. In P_i-bound and IDP- or Ca^{2+}-inhibited structures the hydrophobic gate is fully closed (130).

1.3.6. Dimer interface

Typically, the dimer interface is not counted as one of the functional core regions. This might change in future due its role in mediating intra-subunit communication and recent findings of functional asymmetry in M-PPase catalysis (131,136,137). The dimer interface is formed by residues of outer ring helix 10, 13 and 15 (Figure 1.16) that interact with the opposing subunit *via* hydrogen bonds and hydrophobic interactions (127,128). These interactions are subject to helix motions throughout the catalytic cycle (129). For example, cytoplasmic segments of helix 13 move by up to 8 Å to remain near inner ring helices when substrate binds and relay structural changes between subunits *via* an intricate salt-bridge network consisting of $I/R^{10.33'}$, $R^{13.62}$ and $loop_{5\text{-}6}$ residue $E^{5.71}$ (apostrophe indicates affiliation to the other subunit) (Figure 1.16). The structures of *Tm/Vr*-PPase show that subunit interactions differ between subclasses and/or host organisms. In *Tm/Vr*-PPase:Mg_5IDP, $I/R^{10.33}$ $S/I^{10.41}$, $F/N^{10.42}$, $M/Y^{10.48}$, $V/A^{10.59}$, $A/I^{13.30}$, $A^{13.36}$, $A/M^{13.37}$, $Y^{13.40}$, $Y/W^{13.41}$, $R^{13.62}$, $L/V^{15.44}$ and $M/Q^{15.48}$ mediate intra-subunit

communication in both proteins. Of these, $I^{10.41}$ is the only highly conserved residue (~91% pairwise sequence identity) in the dimer interface of M-PPases. In general, side chain properties such as bulkiness, charge or hydrophobicity are more frequently conserved in this region rather than sequence identity.

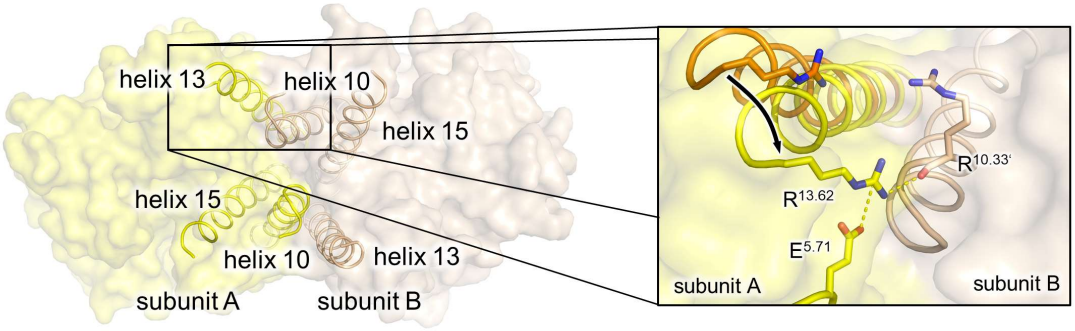


Figure 1.16: Top-down view of the dimer interface in *Vr*-PPase:Mg_5IDP. Subunit A is coloured in yellow and subunit B is coloured in wheat. Dimer interface helices 10, 13 and 15 are annotated in the left panel. The black box shows a close-up view of the intricate salt network comprising $I/R^{10.33'}$, $R^{13.62}$ and $E^{5.71}$ that participates in relaying information regarding structural rearrangements between subunits. The black arrow indicates a major conformational change at helix 13 (*Vr*-PPase:Mg_5IDP in yellow, *Vr*-PPase:Mg_2P_i in orange) that occurs upon substrate binding and reorientation of inner ring helices.

Mutations at various residues at the dimer interface inactivate protein or uncouple ion-translocation and PP_i hydrolysis (121), which further supports the idea that the dimer interface constitutes another functional core region in M-PPases (1.5.2). This would also explain why M-PPases are functional homodimers even though established catalytic key regions are all contained within a single subunit (1.5.2).

1.4. Evolution of M-PPases

The discovery of M-PPases in all three kingdoms of life, and the variety of enzymes differing in ion-pumping specificity, K^+ requirement or Na^+ regulation, indicate an evolutionarily ancient origin (91). PP_i is thought to have been an important energy carrier early in evolution before the rise of adenosine triphosphate (ATP) as the primary energy source (138). The utilisation of energy stored in phosphoanhydride bonds for establishing electrochemical gradients across membranes might have been of even greater importance for early life forms than it is nowadays (138). M-PPases evolved through gene triplication resulting in three conserved splayed bundles of helices 3-6, 9-12 and 13-16 that arrange with 3-fold symmetry perpendicular to the membrane plane and only differ by an rmsd/C_α of 2.1-2.9 Å when superimposed on each other (127). Consistent with studies addressing the evolutionary origin of F- and V-type ATPases, Na^+-PPases are considered to be the predecessor of H^+-PPase and Na^+,H^+-PPases (126,139,140). There is a simple explanation for this: membranes are less leaky to Na^+ than H^+ and therefore Na^+-pumps probably evolved first. Indeed, phylogenetic studies of several PP_i-energised membrane-bound ion pumps revealed the rise of H^+- and Na^+,H^+-PPases

from several independent phylogenetic clades, whereas Na^+-PPases are all found within a single monophyletic clade (91,92). This also supports the idea that Na^+-PPases evolved first.

1.4.1. Potassium (in)dependence

In 1981, a subclass of H^+-PPases have been identified that do not require K^+ for maximal catalytic activity (93). Sequence alignments of H^+-PPases revealed two key amino acid substitutions involved in the evolution of K^+-independence: $A^{12.46}K$ and $G/A^{12.49}T$ more than 20 years later (141). A recent systematic analysis of the $K^+/K^{12.46}$ cationic centre across all M-PPase subfamilies identified $A/K^{12.46}$ as the molecular switch determining K^+-dependence (136). It was shown that $K^{12.46}$ likely masks a conserved K^+ binding site that, when unmasked by the $K^{12.46}A$ mutation, confers partial K^+-dependence on the enzyme (136). However, no structural data of a K^+-independent H^+-PPases supporting this hypothesis are available to date. Moreover, the $A^{12.46}K$ and $G/A^{12.49}T$ change are tightly coupled but there is no functional role ascribed to $G/A/T^{12.49}$ yet. It has been reported that $G/A^{12.49}T$ single variants of K^+-dependent H^+-PPases demonstrate a 3-fold reduced affinity for K^+, but it remains unclear how changes at this position affect the cationic centre ~10 Å away (141).

1.4.2. Ion pumping selectivity

In contrast to K^+-dependence, M-PPase sequence alignments do not reveal conserved residue patterns that reflect experimentally confirmed ion selectivity across *all* M-PPase subclasses. Our current model of ion selectivity is based on the position of a semi-conserved glutamate at the ion gate ($E^{6.53/57}$) (1.3.4). If located one helix turn down ($E^{6.57}$), a Na^+-binding site is destroyed due to the reorientation of $K^{16.50}$, while enabling full protonation of the semi-conserved glutamate in question (Figure 1.14). Evolutionarily, the transition from Na^+ to H^+ pumping simply required repositioning of a single residue without the need of a mechanistic change (129). This supports the idea that small changes in protein sequence likely account for subclass differences as sequence conversation between them is high (95).

However, the current model of ion-selectivity can only explain the transition from K^+-dependent H^+-PPase to K^+-dependent Na^+-PPases and fails when extended to K^+-independent H^+-PPases or K^+-dependent Na^+,H^+-PPases. In the latter two subclasses the molecular basis for H^+-pumping remains unclear as the semi-conserved glutamate did not shift from its position in their evolutionary predecessor, the K^+-dependent Na^+-PPases ($E^{6.53}$), as expected for H^+-PPases. Therefore, the semi-conserved glutamate position cannot be the sole determinant of ion selectivity in M-PPases, which is yet to be uncovered. This is also illustrated by mutational studies that failed to introduce a change in ion selectivity by mirroring the primary ion gate structure of a K^+-dependent

Na^+-PPase in a K^+-dependent H^+-PPase ($E^{6.57}G$ + $G^{6.53}E$) (126). Nevertheless, the semi-conserved glutamate is crucial for structural and functional integrity (1.3.4). Removal of its negative charge was shown to either abolish M-PPase expression ($E^{6.53}S$), uncouple PP_i hydrolysis from ion transport ($E^{6.57}S/Q$) or inactivate the protein entirely ($E^{6.57}A$) (126,130).

1.5. Catalytic cycle

A set of structural studies explored the conformational space of M-PPases by using various ligands and inhibitors to stabilise or trap different protein states throughout the catalytic cycle for X-ray crystallography (127–129,131). This was complemented by electrometric studies (129), kinetic data (131,136) and molecular dynamics (MD) simulations (142) and yielded many mechanistic insights.

The resting state M-PPase structure was obtained by Ca^{2+} inhibition and is characterised by a large active site volume (~3,400 $Å^3$), twice that of IDP-bound structures (129). Atomistic MD simulations showed that the resting state is also of a more dynamic nature than IDP-bound protein, particularly loop regions (142). Indeed, there is no electron density for most cytoplasmic loops when Ca^{2+} is bound to stabilise the resting state (127). The active site remains accessible from the cytoplasm until substrate binds and $loop_{5-6}$ as well as $loop_{13-14}$ become ordered and fold over it (Figure 1.17, step 1). Sealing of the active site is stabilised by water-mediated interaction between $loop_{5-6}$ residues and the metal-substrate complex (129). This also supports binding of the correct substrate as steric hindrance prevents active site closure when bigger molecules such as polyphosphate, nucleoside di- and triphosphates or some bisphosphonates like etidronate occupy the active site (129). $Loop_{13-14}$ folds on top of $loop_{5-6}$ to further secure the binding of correct substrate (129).

The resting state stores strain energy by bending of helix 6 around $D^{6.43}$ (~15° from linear), which is stabilised by the $D^{6.43}$-$K^{12.50}$ ion pair (129). Upon loop closure helix 12 shifts towards the periplasmic/luminal side by ~2 Å, breaking the $D^{6.43}$-$K^{12.50}$ interaction and leading to the release of strain energy by the straitening of helix 6 (129). As a result, $D^{6.43}$ moves by 3-4 Å and now coordinates nucleophilic water together with $D^{16.39}$, which promotes the abstraction of a H^+ for nucleophile generation and subsequent attack on PP_i (Figure 1.17, step 2). Therefore, it was proposed that the downward motion of helix 12 resembles the springing of a molecular mousetrap, which is triggered by substrate binding to the active site and ultimately leads to PP_i hydrolysis (129). In line with the key role ascribed to the $K^{12.50}$-$D^{6.43}$ ion pair for nucleophile activation and conformational changes of inner ring helices, the mutation of $K^{12.50}$ in *Vr*-PPase abolishes both, hydrolysis and H^+-translocation (143). Similarly, mutations hampering the downward motion of helix 12 at a set of conserved and flat hydrophobic residues ($A^{12.53}$, $I^{12.54}$, $A^{12.57}$,

A/I$^{12.58}$) acting as molecular lubricant either uncouple PP_i hydrolysis from H^+-pumping or abolish expression entirely (132).

Mutations hindering the downward motion of helix 12 also affect ion pumping because springing of the mousetrap also induces structural changes at the ion gate such as the corkscrew motion of helix 16 around K$^{16.50}$, which facilitates the localisation of Na^+ or H^+ (Figure 1.17, step 2). It was proposed that ion release into the periplasm/lumen may be triggered by the extended downward motion of helix 12 to momentarily open the exit channel (Figure 1.17, step 3), but there is no data supporting this yet (129). The active site can only reopen after PP_i hydrolysis and pumping. This likely induces structural changes that decrease the affinity for product, which then exits stepwise (129). Once completed, the enzyme is returned into the resting state and ready for another cycle of catalysis (Figure 1.17, step 4-5).

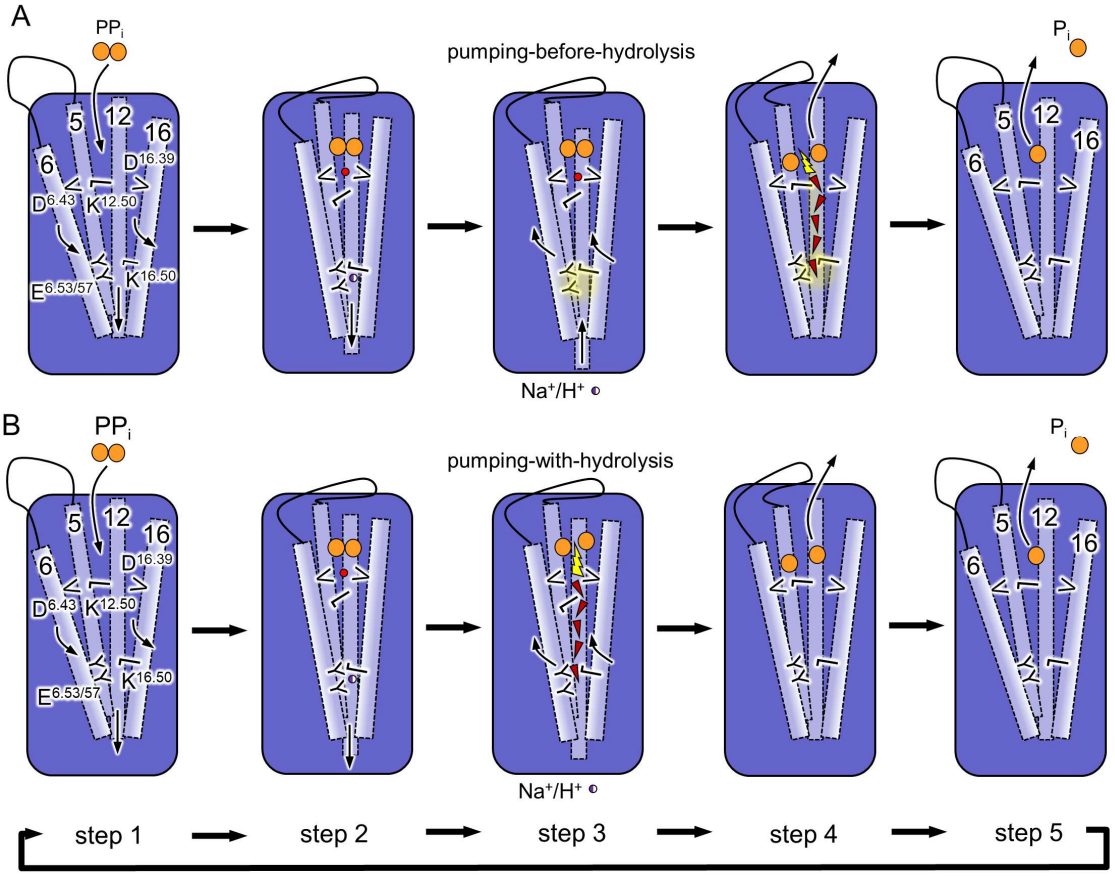


Figure 1.17: Schematic representation of the catalytic cycle and energy coupling in M-PPases. Key residues and helixes are labelled in the first panel. Only one subunit is shown for clarity. Major conformational changes are highlighted by black arrows. Yellow lightning symbol indicates hydrolysis of PP_i and a chain of red arrows represents the ion translocation pathway in the coupling funnel, which connects hydrolysis at the active site to ion pumping *via* the ion gate and exit channel. (A) "Binding change" type energy coupling in which pumping precedes hydrolysis. A yellow glow indicates an overall negative charge that drives the abstraction of a H^+ for nucleophile generation after ion translocation. (B) "Direct coupling" type mechanism in which pumping and hydrolysis take place synchronised. The H^+ originating from PP_i hydrolysis is pumped in subsequent cycles or pushes a Na^+ that is bound at the ion gate into the exit channel ("billiard-type" mechanism) for it to be released to the periplasm/lumen.

1.5.1. Coupling

The chronological order of PP_i hydrolysis and ion pumping in M-PPases is vehemently discussed (129,144). The two opposing mechanisms of energy coupling either postulate "pumping-before-hydrolysis" (Figure 1.17A) or "pumping-with-hydrolysis" (Figure 1.17B). The "pumping-with-hydrolysis" model, also called "direct-coupling", postulates that PP_i hydrolysis and ion pumping occur simultaneously and that the H^+ release from nucleophilic water during PP_i hydrolysis is the one pumped after n cycles (Figure 1.17B), where n is the number of downstream ion binding sites (144). This was extended by a "billiard-type" mechanism to explain Na^+-transport in which the generated H^+ pushes Na^+ in the exit channel for pumping (145). In contrast, the "pumping-before-hydrolysis" model, also called "binding-change", favours a mechanism in which ion pumping precedes hydrolysis (Figure 1.17A) and is triggered by the closure of the active site (129). The transported ion may originate from the medium or preceding hydrolysis events and can explain both, H^+- and Na^+-pumping. The overall negative charge at the ion gate that results from "pumping-before-hydrolysis" could then induce the downward motion of helix 12 and associated events, such as the $K^{12.50}$-$D^{6.43}$ ion-pair breakage and nucleophile activation by $D^{6.43}$ and $D^{16.39}$ (1.5). The negative charge would also promote the abstraction of a H^+ from nucleophilic water at the active site and thereby drive the hydrolysis of PP_i. The generated H^+ would then enter the Grotthus chain to reset the ion gate (129). Alternatively, nucleophile generation may follow a general acid/base-type catalysis as in aspartyl protases (146). This requires the protonation of either $D^{6.43}$ or $D^{16.39}$ around pH ~7 at which M-PPases operate; unlikely but not impossible ($pK_{a,D}$: 3.65). However, an acid/base-type catalysis is not compatible with "direct coupling" as it does not release a H^+ that could be pumped in subsequent cycles.

The ongoing debate about energy coupling between hydrolysis and ion translocation across the membrane in M-PPases reflects a similar discussion for F_1F_0-ATPases in the 1970s, which are functionally but not structurally analogous to M-PPases. Peter Mitchel first proposed the "direct-coupling" model where the energy transfer between substrate and H^+ are directly linked, *i.e.* H^+ takes part in ADP phosphorylation during ATP synthesis at the active site (147). Paul Boyer supported a "binding-change" or "rotational mechanism" in which catalysis depends on allosterically generated conformational strain by substrate hydrolysis at F_1 or H^+ flow through F_0, depending on its direction of work (ATP hydrolysis *versus* synthesis) (148). This debate was resolved in favour of the "binding-change" model, which aligns with experimentally determined structures. Contrary to F_1F_0-ATPases, the available structural data on M-PPases cannot discriminate between the "direct-coupling" and "binding-change" mechanism as the captured conformations mimic reaction intermediates that can describe both mechanisms (144).

Therefore, Li and co-workers investigated the chronological order of events in electrometric studies to complement the available structural data and resolve the debate of energy coupling (129). Liposomes containing K^+-dependent H^+-PPase from *V. radiata* (*Vr*-PPase) were immobilised on membrane-coated gold chips to measure the ion flow (currents) across membranes upon applying substrate (PP_i), product (P_i) or non-hydrolysable substrate analogues (IDP, etidronate) (129). Besides PP_i, only IDP induced a current (Figure 1.18A-B). This was associated to H^+ translocation in gramicidin and carbonyl cyanide m-chlorophenyl hydrazone (CCCP) control experiments (Figure 1.18B-C) (129,142). Etidronate and P_i did not induce a comparable signal (Figure 1.18C). Based on these findings the authors drew two conclusions: The first one is that hydrolysis is not required for ion transport. The second one is that ion transport depends on the closure of the active site, only possible with PP_i or very close analogues, but not with etidronate as it is too bulky (129). This interpretation clearly favours a "binding change" mechanism in which ion pumping precedes hydrolysis.

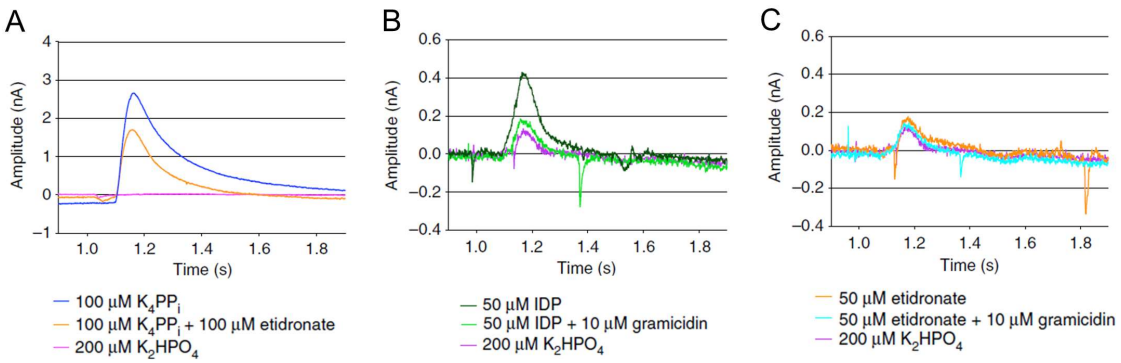


Figure 1.18: Electrometric sensor traces of the Nanion SURFE²R N1 machine measuring H^+-pumping in *Vr*-PPase proteoliposomes. (A) PP_i induced signal, P_i control and effect of etidronate addition. (B) IDP induced signal, P_i control and effect of gramicidin addition. (C) Etidronate induced signal, P_i control and effect of gramicidin addition. Adapted from Li *et al*. (129).

Baykov, an M-PPase enzymologist, later opposed this (144) since PP_i-binding led to a 10-fold bigger signal compared to IDP despite similar binding affinities (Figure 1.18A-B). The signal is generated within 0.1 seconds of ligand addition, a timeframe he ascribes to a single turnover event based on substrate turnover measurements in external studies (144). Baykov suggests that the PP_i-signal is indicative of a full pumping event, whereas substrate-analogues only lead to a partial membrane crossing of ions, for example PP_i pulling in Mg^{2+} from the medium into the active site as Mg_2PP_i. Alternatively, he ascribes the lower signal induced by non-hydrolysable ligands to the reorientation of charged amino acids in functional core regions or simply to the ionic nature of the substrate-analogues. This interpretation supports a "direct-coupling" mechanism in which hydrolysis is key to, and synchronised with, ion transport.

However, the argumentation ignores the fact that etidronate and P_i did not induce a comparable signal in electrometric studies although they both bind with Mg^{2+} to the active

site, are both of ionic nature and both induce structural reorientations of charged residues, just like IDP (Figure 1.18C). Moreover, the 10-fold difference between PP_i and IDP induced signals in electrometric studies could be explained by multiple turnover events, only possible with PP_i. Indeed, complementing hydrolytic activity data obtained from proteoliposomes used in electrometric studies indicates a substrate turnover of >1 (k_{cat}: ~60 s^{-1}) within the timeframe of signal generation (129). Finally, "direct-coupling" alone does not provide a unifying mechanism explaining Na^+ transport, for which the "billiard-type" mechanism is required in addition. So far, the only physiologically relevant Na^+-binding site that has been identified is at the ion gate (90). A H^+ originating from the hydrolysis event would need to travel two-thirds through membrane-spanning protein regions and back just to push Na^+ through. Nevertheless, the idea of "direct-coupling" sticks until today.

This dispute may be resolved in future with the availability of brighter neutron sources, which will potentially enable the mapping of H^+ positions in crystal structures of more challenging targets than possible nowadays.

1.5.2. Asymmetry and inter-subunit communication

In the 1990s, radiation inactivation experiments revealed subunit interdependence of M-PPases (149–151), but only recently reports of allostery resparked interest in the fact that M-PPases are functional dimers even though all catalytic core regions are contained within a single subunit (131,136,137,152,153). The data of two independent kinetic studies revealed substrate inhibition in a set of different M-PPases (131,136). As all available structural data shows no evidence for alternative binding sites, nor is there any space for PP_i binding away from the active site, the observed inhibitory effects must be conferred by substrate binding to the second subunit.

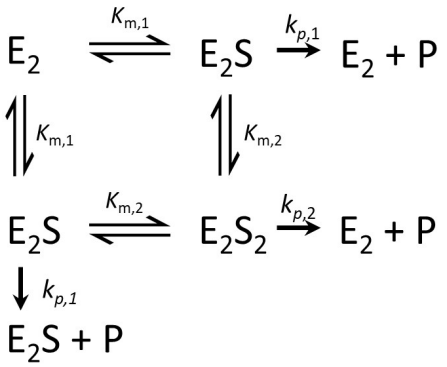


Figure 1.19: Kinetic model of allosteric Interaction between active sites of M-PPase subunits.

Table 1.3: Exemplary kinetic parameters of PP_i binding and hydrolysis for two different M-PPases.

Parameters	***Tm*-PPase**	***Bv*-PPase**
Host organism	*Thermotoga maritima*	*Bacteroides vulgatus*
Classification	K^+-dependent Na^+-PPase	K^+-dependent true Na^+·H^+-PPase
Binding cooperativity	positive	negative
Kinetic parameters*		
$K_{m,1}$ (µM)	666 ± 11	23 ± 2
$K_{m,2}$ (µM)	81 ± 19	800 ± 4
$k_{p,1}$ (µmol P_i mg^{-1} min^{-1})	235 ± 28	0.28 ± 0.01
$k_{p,2}$ (µmol P_i mg^{-1} min^{-1})	12 ± 1	0.060± 0.01

* Determined in presence of 50 mM K^+, 10 mM Na^+ and 5 mM free Mg^{2+}

Figure 1.19 depicts a kinetic model that is based on these findings and assumes an allosteric interaction between the two subunits in M-PPases (136). E_2 represents dimeric

enzyme, S is substrate (Mg_2PP_i), P is product (P_i), K_m denotes the microscopic Michaelis-Menten constant and k_p gives the per-site maximum velocity of substrate hydrolysis.

It was demonstrated that occupation of both active sites leads to a 3 to 20-fold decrease in the maximal velocity of substrate hydrolysis in M-PPases (131,136). Interestingly, PP_i binding to the first active site affected binding to the neighbouring subunit differently (131,136). Artukka and co-workers reported a decrease in PP_i affinity at the second subunit, indicating negative cooperativity, whereas Vidilaseris and co-workers saw an increase in PP_i affinity at the second subunit, indicating positive cooperativity (131,136). Table 1.3 shows the kinetic parameters of two M-PPases (131,136), one exemplary for positive substrate binding cooperativity ($K_{m,1} > K_{m,2}$) and one exemplary for negative substrate cooperativity ($K_{m,1} < K_{m,2}$). Since then, more studies have reported ambiguities with respect to the cooperativity of ligand binding to M-PPase active sites (137,152,153). Taken all available data together, the nature of binding cooperativity appears to differ on protein-by-protein basis as there is no evident link to subclass affiliation (*e.g.* K^+-dependent *versus* K^+-independent M-PPases), ligand type (*e.g.* PP_i-binding *versus* inhibitor-binding) or, as initially thought, solubilisation method (*e.g.* detergent-based *versus* lipid-based). The mechanistic details of this are unclear.

Nevertheless, occupation of both active sites consistently led to reduced hydrolysis rates ($k_{p,1}<k_{p,2}$) and subunit interdependence appears to be tweaked by the precise residue configuration at the active site, not just by its occupation status (131,136,137,152,153). For example, in the K^+-dependent H^+-PPase from *Desulfitobacterium hafniensee* (*D. hafniensee*), the affinity of PP_i binding to the second subunit was an order of magnitude greater when the first subunit was occupied by inhibitor instead of substrate, indicating that substrate and inhibitor induce non-equivalent structural changes at the respective other subunit (137). Similarly, the authentic state of the $K^+/K^{12.46}$ cationic centre and not its occupation status was as shown to be crucial for substrate inhibition in a range of mesophilic M-PPases. In the absence of K^+, substrate inhibition is abolished in K^+-dependent H^+-PPases (136). Although Na^+ can maintain hydrolytic activity in K^+-free medium by occupying the K^+ binding site, it was unable to reinstate the inhibitory effects of excess substrate (136). Similarly, $A^{12.46}K$ variants failed to confer substrate inhibition but did facilitate K^+-independence (136).

The first structural evidence of asymmetry in M-PPases was obtained in presence of the non-phosphorus allosteric inhibitor N-[(2-amino-6-benzothiazolyl)methyl]-1H-indole-2-carboxamide (ATC) of *Tm*-PPase (131). ATC was found to bind as a dimer to periplasmic loop regions of one of the two subunits, where it acts as a hydrophobic clamp and keeps $loop_{6\text{-}7}$, $loop_{8\text{-}9}$ and $loop_{12\text{-}13}$ together. This locks the exit channel in a close state (131).

ATC_2 only binds when substrate (analogue) is present at the active site as it captures a conformation that is induced upon the downward motion of helix 12, subsequent $loop_{12\text{-}13}$ reorientation towards the dimer interface and reorientation of helix 13 (131). This links substrate-binding at the active site to intra-subunit communication at the dimer interface. If ATC_2 is bound, full motion required for catalysis is restricted (Figure 1.20E) but binding of substrate (analogue) to the active site is still communicated *via* helix 13 (dimer interface) into the other subunit where it exhibits conformational changes affecting the binding cooperativity positively (Figure 1.20F).

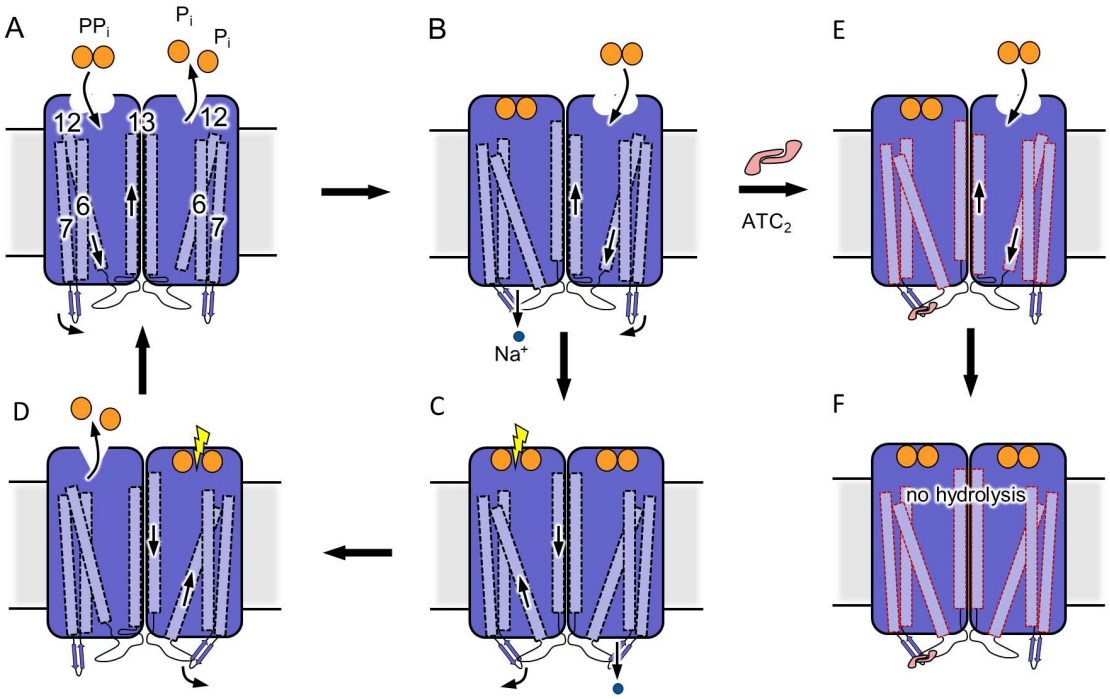


Figure 1.20: Schematic representation of functional asymmetry and the „half-of-the-sites“ reactivity model in *Tm*-PPase. Key helixes are labelled in the first panel and major conformational changes are indicated by black arrows. (A) Resting-state enzyme with active sites of subunit A and subunit B unoccupied. (B) PP_i binds to subunit A where it triggers structural changes that lead to Na^+-pumping (“binding change” type model) and increase the PP_i binding affinity of subunit B. (C) Na^+ pumping in subunit A now triggers PP_i hydrolysis here. PP_i binding to subunit B leads to structural changes that result in Na^+-pumping and alter the binding affinity for product at subunit A. (D) Product is released in subunit A and Na^+-pumping in subunit B now leads to PP_i hydrolysis here. The resting state is reinstalled in the next step by product release at subunit B. (E) ATC_2 binds to periplasmic/lumenal loops of subunit A and locks the enzyme in a substrate-bound conformation with the exit channel closed and helix 12 in a “down” state. (F) ATC_2 binding restricts the full motions required for hydrolysis. The structural changes triggered by PP_i binding in subunit A increase the affinity for PP_i at subunit B. Upon PP_i binding to subunit B structural changes remove the asymmetry in unrestricted protein parts.

Both the kinetic and the structural data suggest a *modus operandi* in which only one subunit is predominantly active at any given time („half-of-the-sites“ reactivity model) and structural changes are relayed over the dimer interface into the respective other subunit (Figure 1.20A-D). The simultaneous occupation of both active sites likely removes asymmetry between subunits and ultimately hinders catalysis. Consequently, the occupation of both active sites in crystal structures might be a crystallographic artefact

of using non-hydrolysable inhibitors or excess ligand concentrations for co-crystallisation and therefore mechanistically irrelevant. Vidilaseris and co-workers proposed a model for functional asymmetry in *Tm*-PPase (131), in which PP_i binding to subunit A (Figure 1.20A) induces structural changes in subunit B that prime the active site for the arrival of a second PP_i molecule and promote its binding (Figure 1.20B) in accordance with kinetic data showing positive cooperativity (Table 1.3) (131). Upon binding of a second PP_i molecule to subunit B, PP_i in subunit A gets hydrolysed (Figure 1.20C) and product is released (Figure 1.20D). This then allows hydrolysis of the second PP_i, product release in subunit B and enzyme returning to a resting state (Figure 1.20A). Depending on the model of energy coupling, the ion transport occurs either at the substrate binding step and drives hydrolysis ("binding-change"), which is favoured by Vidilaseris and co-worker and depicted in Figure 1.20, or upon PP_i hydrolysis ("direct-coupling") (1.5.1).

Notably, the „half-of-the-sites" reactivity model is also consistent with the idea that allosterically induced asymmetry between subunits promotes the selective binding of either Na^+ or H^+ in Na^+,H^+-PPases, explaining dual-pumping in proteins with only a single ion-translocation pathway per subunit (91). To date, there is not structural data available for Na^+,H^+-PPases to investigate the putative role of functional asymmetry in ion selectivity. For a detailed mechanistic insight into this complex topic, time-resolved biophysical experiments probing physiological enzyme states and complementary MD simulations are required.

1.6. M-PPases in human health and global food security

The ability of various plants, prokaryotes and parasites to utilise the PP_i pool as an additional energy source is crucial for surviving low-energy and high-stress conditions by establishing electrochemical gradients across membranes (1.2.3). This makes M-PPases a valuable target for structural and functional studies, which could be of benefit in the fight against existing and emerging challenges to human health and global food security. For example, climate change is already causing more frequent and extreme droughts and heatwaves, reducing global crop yields by 3.2-7.4% (depending on the crop type) per degree rise in global mean temperature (155–159). The overexpression of M-PPases has been shown to improve tolerance to drought in various transgenic plants and could therefore offer a possible solution to this growing problem (104–112). Climate change also facilitates the northwards spread of insect vectors for viruses and parasites such as *Plasmodium falciparum* (*P. falciparum*), putting millions of new people at risk of life-threatening tropical diseases in the coming decades (160,161).

While enhancing or introducing M-PPase expression confers stress resistance in plants (104–112), the inverse approach is equally promising to impair cell homeostasis

of M-PPase carrying pathogens, especially under stress conditions (162–164). Relevant pathogens include a range of protozoan parasites causing severe diseases such as malaria (*Plasmodium* ssp.), leishmaniasis (*Leishmania spp*.), trypanosomiasis (*Trypanosome spp.*) and toxoplasmosis (*Toxoplasma gondii*) as well as several organisms of the bacterial genus of which *Bacteroides spp*. are particularly troubling (96). This is because *Bacteroides spp*. display very high antibiotic resistance rates and can cause significant pathology, for example bacteraemia and abscess formation in multiple body site including the brain (165). Depending on the availability of treatment options, brain abscesses caused by *Bacteroides spp.* are associated with mortality rates of 19-60% (165). So far, studies investigating the potential of M-PPases as drug target focused on relevant parasitic pathogens. In *P. falciparum*, M-PPase knockout impaired acidocalcisome acidification, which is thought to result in reduced adaptability to osmotic pressure (162). In line with this assumption, similar studies led to reduced virulence and lower in vitro asexual blood stage growth of *Trypanosoma brucei* and *Toxoplasma gondii* (163,164). These findings suggests that targeting M-PPases is a worthwhile strategy to treat a range of serious diseases associated with a high prevalence and risk of fatality, particularly in the light of growing antimicrobial resistance (166). This is further supported by the fact that there are no human M-PPase analogues and therefore off-target binding that can cause severe side effects is less likely to occur when targeting M-PPases.

1.6.1. Recent advances in drug discovery

To date, there are no drugs tested or approved that impair M-PPases function in pathogens. The specific targeting of these proteins has proven difficult as phosphorus inhibitors are often recognised by unrelated structural motifs (167). Recently, a set of non-phosphorus M-PPases inhibitors have been found in fragment screening campaigns (131,167,168). Their identification was based on hydrolytic activity measurements in a 96-well format using thermostable *Tm*-PPase as model target to increase throughput (169). Initial fragment hits were developed by directed chemical exploration to increase their potency, which resulted in improved IC_{50} values (concentration at which activity is halved) in the low µM range (131,167,168). The *Tm*-PPase test system proved suitable for high-throughput screening in an academic setting but did not translate well to pharmaceutically relevant targets. For example, one of the best hits identified in terms of its IC_{50} (1.7 µM), ATC_2, did not exhibit any anti-plasmodial activity in *P. falciparum* survival assays (131). Sequence analysis revealed a high degree of difference at ATC-binding periplasmic loop regions of *Tm*-PPase compared to homologous proteins from parasites, *i.e.* the absence of $loop_{6\text{-}7}$ in the latter (131). This explains the lack of efficacy for ATC_2 in *P. falciparum* and highlights a weakness of the *Tm*-PPase test system. To date, screening continues and

more lead molecules such as isoxazole-based compounds have been identified with promising IC_{50} values for *Tm*-PPase. Some have potential for further development into therapeutic molecules as they show experimentally verified inhibitory effects on *P. falciparum* growth (167,168).

1.7. Aims and objectives of this project

The aim of my thesis was 2-fold. As part of the "Rationalising Membrane Protein Crystallisation" (RAMP) innovative training network (ITN) I aimed to drive the development of innovative and easy-to-use tools that address major bottlenecks of membrane protein structure determination, in particular protein instability. At the same time my project was centred on M-PPases, of which a few can already be purified in sufficient quantities and quality for crystallographic structure determination. Therefore, I used M-PPases as a test system to investigate novel tools for membrane protein stabilisation. This was done by either predicting stabilising point mutations with sufficient accuracy to speed up and simplify the generation of stable constructs or by re-lipidating detergent-solubilised membrane-proteins to reinstate specific protein-lipid interactions that are key for functional and structural integrity in a collaborative effort.

Moreover, I used the newly developed tools to aid the functional and structural investigation of M-PPase subgroups for which no structural data is available yet and answer open mechanistic questions (1.3-1.5), for example: What is the structural basis for K^+-independence, what is the role of the A/$G^{12.49}$K coupled change in K^+-independent M-PPases, how is the K^+/$K^{12.46}$ cationic centre linked to substrate inhibition and intra-subunit communication, what is the molecular mechanism underlying functional asymmetry, how is PP_i hydrolysis and ion pumping coupled and what defines ion-pumping selectivity in K^+-independent H^+-PPases and K^+-dependent Na^+,H^+-PPases? The structural studies focused on static and time-resolved X-ray crystallography of a K^+-independent H^+-PPase (*Pa*-PPase), a K^+-dependent Na^+,H^+-PPase (*Cl*-PPase) and a K^+-dependent Na^+-PPase (*Tm*-PPase).

The findings of my project have the potential to greatly support the scientific community that investigate challenging and high-impact target proteins by providing tools for their stabilisation. Moreover, functional and structural characterisation of different M-PPases will provide new mechanistic insights that could ultimately benefit drug discovery efforts and lead to the development of therapeutic molecules treating a wide range of serious and "difficult-to-treat" diseases.

Chapter 2. Material and Methods

2.1. Material

2.1.1. Chemicals

All chemicals used were purchased from Sigma-Aldrich, Thermo Fisher Scientific, Melford Laboratories, Alfa Aesar, VWR International, Honeywell Fluka™, Avanti or Anatrace unless stated otherwise.

2.1.2. Buffers and Solutions

All buffers and solutions used were prepared using Milli-Q® water and filtered with 0.22 µm or 0.45 µm membrane filters (GE Healthcare Life Sciences). Buffers prepared for use in Fast Protein Liquid Chromatography (FPLC) systems were additionally degassed prior to use. Table 2.1 lists commonly used buffers and solutions, which are referred to in this thesis. All other buffers used are described in the text.

Table 2.1: List of commonly used buffers and solutions.

Buffer and solutions	Composition
5x reducing protein loading buffer	10% *w/v* sodium dodecyl sulphate (SDS), 500 mM dithiothreitol (DTT), 50% *v/v* glycerol, 500 mM Tris-HCl pH 6.8, 0.05% *w/v* bromophenol blue
SDS-PAGE running buffer	25 mM Tris-HCl pH8.3, 192 mM glycine, 0.1% *w/v* SDS
Tris-acetate buffer (TAE)	40 mM Tris-acetate pH 8.5, 1 mM ethylenediaminetetra-acetic acid (EDTA)
Tris-buffered saline (TBS)	20 mM Tris-HCl p 7.6, 150 mM NaCl
Transformation buffer 1 (TBF1)	30 mM KOAc, 100 mM RbCl, 50 mM $MnCl_2$, 10 mM $CaCl_2$
Transformation buffer 2 (TBF2)	10 mM MOPS, 10 mM TbCl, 75 mM $CaCl_2$, 15% *v/v* glycerol
Phosphate-buffered saline (PBS)	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4

2.1.3. Media and agar

All media and agar used were prepared using Milli-Q® water and sterilised by autoclaving. Heat sensitive components (*e.g.* glucose, galactose, isopropyl-β-D-thiogalactoside (IPTG), antibiotic stocks, amino acid mixes) were sterilised by 0.22 µm filtration using membrane filters and added after autoclaving. Table 2.2 lists commonly used media that are referred to in this thesis. Plates were prepared by adding 1.5% *w/v* agar. All other media used are described in the text.

Table 2.2: List of commonly used media.

Media	Composition
Lysogeny broth (LB)	0.5% w/v yeast extract, 1% *w/v* tryptone,1% *w/v* NaCl
Terrific broth (TB)	1.2% w/v yeast extract, 2.4% *w/v* tryptone, 0.4% *v/v* glycerol, 17 mM KH_2PO_4, 72 mM K_2HPO_4
Super optimal broth catabolite repression (SOC)	0.5% w/v yeast extract, 2% *w/v* tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose
Yeast extract peptone dextrose media (YPD)	1% w/v yeast extract, 2% *w/v* peptone, 2% *w/v* glucose

2.1.4. Primers

All primers used were synthesised by Eurofins Genomics and desalted. A list of primers used in this project can be found in Table S2 (Appendix).

2.1.5. Constructs

The sequences of all constructs are provided in FASTA format in Table S3 (Appendix). More detailed information such as the molecular weight, extinction coefficient, tryptophan content and isoelectric point (pI) for each construct are given in Table S4 (Appendix).

2.2. General methods of molecular biology

2.2.1. *Escherichia coli* transformations

All transformations were done following the standard heat shock protocol with in-house prepared RbCl competent *Escherichia coli (E. coli)* strains. In brief, glycerol stocks of *E. coli* were plated on LB agar. A single colony was used to inoculate a starter culture in 5 mL SOC media for an overnight incubation at 37 °C. The starter culture was then 1:100 diluted in SOC media, shaking at 200 rpm, to yield a growth culture. The growth culture was cultivated at 37 °C, shaking at 200 rpm, until the cells reached an optical density (OD) of 0.6 at 600 nm. Cells were harvested by centrifugation (3,000 *x*g, 4 °C, 5 minutes), washed in cold 100 mL TBF1 buffer and resuspended in cold 10 mL TBF2 buffer per 200 mL of cell culture. After 15 minutes on ice, the cells were either used for transformation or flash frozen in liquid nitrogen for later use. For transformations, ~100 ng of plasmid was added to 50 µL competent cells. After 20 minutes on ice, the cells were heat shocked at 42 °C for 45 seconds and returned to ice for another 2 minutes. 1 mL of sterile LB or SOC media was added and the cells were incubated for 1 the cells were pelleted (3,000 *x*g, 4 °C, 5 minutes), resuspended in sterile media and plated out on selection plates. Colonies usually formed after 12 hours at 37 °C.

2.2.2. *Saccharomyces cerevisiae* transformations

All transformations were done following the standard heat shock protocol for yeast. In brief, *Saccharomyces cerevisiae* (*S. cerevisiae*) strains were incubated at 30 °C for 72 hours on YPD pates. A single colony was used to inoculate 5 mL YPD starter culture, of which 1 mL was used per transformation after an overnight incubation at 30 °C, shaking at 200 rpm. Cells were harvested by centrifugation (3,000 *x*g, 4 °C, 5 minutes) and washed in MilliQ water. Afterwards, cells were resuspended in 10 µL yeast transformation buffer (10 mM LiAc, 33.3% *w/v* PEG3350, 100 µg denatured fish sperm DNA, 100-500 ng plasmid) and heat shocked at 42 °C for 40 minutes to trigger DNA uptake. Finally, cells were collected by centrifugation (3,000 *x*g, 4 °C, 5 minutes), resuspended in sterile MilliQ and plated out on selection plates for incubation at 30 °C until colonies formed (~72 hours).

2.2.3. Agarose gel electrophoresis

PCR products were visualised using 0.8-1% *w/v* agarose gels prepared in 1x TAE buffer with 0.5x SYBR™ Safe DNA gel stain (Invitrogen) added. 2 µL of PCR product were loaded alongside a GeneRuler DNA ladder mix (Thermo Fisher Scientific). Prior to loading, all samples were supplemented with 1x purple gel loading dye (New England BioLabs). Gels were run were run in 1x TAE buffer at a constant 100 V and visualised using a G:BOX (Syngene).

2.2.4. Polyacrylamide gel electrophoresis

Samples for denaturing sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) were treated with 5x reducing protein loading buffer and incubated for a minimum of 5 minutes at room temperature before loading into Mini-Protean® TGX (4-20%) precast gels (Bio-Rad) alongside colour prestained protein standard (Broad Range, New England BioLabs) as marker. Proteins were separated by running at a constant voltage of 150 V in SDS-PAGE running buffer and visualised by Coomassie staining (Quick Coomassie Stain, Generon), Western-Blotting (2.2.5) or green fluorescent protein (GFP) fluorescence imaging using a G:BOX (Syngene) with appropriate settings.

2.2.5. Western-Blot

Where required, proteins were transferred onto polyvinylidene fluoride (PVDF) membranes using Trans-Blot® Turbo™ Transfer Packs (Bio-Rad) with a Trans-Blot® Turbo™ Transfer System (Bio-Rad) for Western-Blotting following separation by SDS-PAGE as described in 2.2.4. The PVDF membranes were blocked in 3% *w/v* bovine serum albumin (BSA), which was dissolved in TBS supplemented with 0.1% *v/v* Tween-20 (TBS-T). Blocking took place over the course of 1 hour at room temperature or 12 hours at 4 °C on a roller mixer. Blocked membranes were incubated with either primary antibody or horseradish peroxidase (HRP) conjugated antibody using an appropriate antibody dilution (1:5000 if not stated otherwise) for 1 hour at room temperature on a roller mixer. Excess antibody was removed by three 15-minutes washes with TBS-T at room temperature on a roller mixer. If required, the same procedure was followed for the incubation with an appropriate secondary antibody conjugated to HRP. Afterwards, the membrane was developed with Clarity™ Western ECL Substrate (Bio-Rad) and imaged using a G:BOX (Syngene).

2.2.6. Bradford assay

To determine the total membrane protein concentration Bradford assays were performed. The assay reagent (Thermo Fisher Scientific) was warmed to room temperature before usage. Bovine gamma globulin was used as protein standard at a range of final concentrations between 0.02-2 mg mL^{-1} to obtain a calibration curve that was later used to calculate the sample protein concentration. 20 µL of neat, diluted sample (1:10, 1:100, 1:1000 and 1:5000), standard or buffer control were added to 180 µL of Bradford reagent. The assay was done in transparent round-bottom 96-well microplates. After 20 minutes at room temperature the absorbance at 595 nm was measured using a Spark microplate reader (Tecan). The absorbance readings were inserted into the linear equation of the calibration curve to calculate the protein concentration of the sample.

2.2.7. PCR

To 5 µL of Q5® High-Fidelity 2X master mix (New England BioLabs), 1 ng template plasmid, 0.5 µM forward and 0.5 µM reverse primer (final concentration) were added and the sample was made to a volume of 10 µL using nuclease free water. Inverse PCR was done with back-to-back primers to linearise plasmid or introduce point-mutations. All primer sequences are listed in Table S2 (Appendix). The following program was used for the PCR reactions in a T100 thermocycler (BioRad): initial heat activation at 98 °C for 5 minutes, 30x cycling through 30 seconds of denaturation at 98 °C, 30 seconds of annealing in accordance with the primer melting temperatures as calculated based on their sequence (https://tmcalculator.neb.com/) and 3 minutes of extension at 72 °C, followed by a final extension for 5 minutes at 72 °C and holding at 4 °C. PCR products were visualised on 0.8-1% *w/v* agarose gels (2.2.3).

2.2.8. Quantitative P_i-release assays

The hydrolytic activity of M-PPases was determined by quantifying the P_i release as reduced 12-molybdophosphoric acid (170). The specific reaction conditions, protein and substrate concentrations as well as incubation temperatures and times varied and are described in individual method sections. The following outlines the general workflow, which was identical for all activity measurements.

P_i-release assay solution A (17 mM ascorbic acid in 0.5 M HCl) and B (6 mM ammonium-heptamolybdate) were freshly prepared and stored on ice until used. Additionally, a 30 mg mL^{-1} stock of L-α-lecithin from soybean was prepared by 1-minute intervals of sonication in 20 mM Tris-HCl pH 8.0 and 1 mM DTT, interspersed by cooling on ice until the solution turned clear. Activity measurements on purified samples were done with relipidated protein unless stated otherwise. 10 µL of L-α-lecithin stock was mixed with 5.2 µL of 25% *w/v* n-Dodecyl-β-D-Maltopyranoside (DDM) and incubated at 55 °C for

15 minutes. Afterwards, the mixture was topped up to 25 µL with protein buffer and with purified protein to a final concentration of 12 mg mL^{-1} L-α-lecithin, 600x CMC of DDM and 0.4-3.6 mg mL^{-1} M-PPase. If not stated otherwise, neat protein buffer replaced the protein sample in negative controls. Relipidated samples were diluted 200x in reaction buffer for subsequent activity measurements at a protein or experiment specific temperature. Activity measurements on membranes were done by directly adding resuspended membranes at 2-4 mg mL^{-1} total membrane protein concentration to the reaction mixture instead. Details about incubation temperature, incubation time, protein concentration, reaction conditions and reaction initiation are given in their respective method sections (2.4.5, 2.5.4, 2.6.4, 2.7.3, 2.8.1).

Generally, the reaction mixtures were generally composed of 50-100 mM Tris-HCl pH 7-8, 3-5 mM $MgCl_2$, 0-100 mM KCl, 10-20 mM NaCl and 0-1400 µM Mg_2PP_i. The effective Mg_2PP_i concentration in the reaction condition was approximated as described by Baykov and co-workers (171). P_i calibration curves were prepared by supplementing the final reaction condition with 0-800 µM Na_2HPO_4 instead of M-PPase sample. All reaction components were pre-heated to the incubation temperature prior to use. Substrate hydrolysis was tightly time controlled to keep the incubation period of all samples within each experiment identical. This was done by adding trigger solution (*e.g.* PP_i) to start the reaction every 20 seconds to an 8-strip of 0.2 mL PCR tubes using a multi-channel pipette. Each 8-strip of tubes was incubated in a T100 thermocycler (BioRad) and then transferred stepwise to ice every 20 seconds to stop the reaction. The incubation on ice continued for 10 minutes before 60 µL of a 1:1 mixture of P_i-release assay solution A and B were added in 20 second intervals to form coloured 12-molybdophosphoric acid. The colour formation was allowed to proceed for 10 minutes. Afterwards, excess molybdate was complexed by the stepwise addition of 90 µL of sodium arsenite solution (68 mM tribasic sodium citrate, 154 mM sodium arsenite, 2% *v/v* acetic acid, 0.25% *v/v* Tween-20) to each 8-strip every 20 seconds. Afterwards, the assay was developed for 1 hour at room temperature. P_i amounts were determined based on the absorbance recorded at 860 nm for the standards, blanks and samples using a Spark microplate reader (Tecan). For comparison of activity levels, the specific protein activity defined as the amount of released P_i per minute and amount of M-PPase (µmol P_i mg^{-1} min^{-1}) or the catalytic PP_i turnover (k_{cat}) defined as the number of hydrolysed substrate molecules at each active site per second was calculated. These calculations were based on the linear equation of the P_i calibration curve.

2.2.9. Mass spectrometry

All mass spectrometry data were acquired and processed by Dr. Rachel George of the Mass Spectrometry Facility at the University of Leeds. In brief, protein desalting and

mass analysis were performed by liquid chromatography mass spectrometry (LC-MS) using a nanoAcquity LC system connected to a Xevo G2-S mass spectrometer (Waters Ltd.). Sample at 12-15 µM was loaded onto an Acquity UPLC Protein BEH C4 column (Waters Ltd.) connected to a Acquity UPLC Protein BEH VanGuard pre-column (Waters Ltd.). The BEH C4 column was washed with 10% *v/v* solvent B (1% *v/v* formic acid in acetonitrile) in solvent A (0.1% *v/v* formic acid in water) for 3 minutes at 50 µL /min. Afterwards, bound proteins were eluted using a gradient of 10-95% *v/v* solvent B in A over 10 min at 50 µL/min. The eluant was directed into the mass spectrometer operating in positive time of flight mode (capillary voltage: 3.0 kV, sample cone: 60 V, source offset: 80 V) *via* a Z-spray electrospray source. The source temperature was set at 100 °C and desolvation was performed at 250 °C. Mass calibration was done by a separate injection of 100 fmol μL^{-1} Glu-1-Fibrinopeptide B. Data processing was carried out in MassLynx 4.2.

2.3. Development of a pre-crystallisation lipid screen

2.3.1. Screen design and production

All lipids used in the RAMP lipid screen were purchased from Avanti or Anatrace, where possible already dissolved in chloroform. Lipids supplied as powder were solubilised in chloroform or a chloroform methanol mixture according to their solubility. DPPE (1,2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine) was dissolved in a 9:1 mixture of chloroform and methanol at 5 mg mL^{-1}. POPG (1-palmitoyl-2-oleoyl-*sn*-glycero-3-phospho-1'-*rac*-glycerol) was dissolved in a 5:1 mixture of chloroform and methanol at 10 mg mL^{-1}. All other lipids were prepared in chloroform at a stock concentration of 10 mg mL^{-1}. The equivalent of 0.3 mg of each lipid were then transferred into a round-bottom glass-coated 96-well WebSeal Plate+ (Thermo Fisher Scientific). For lipid mixtures the total transferred lipid amount was 0.3 mg with individual lipid amounts dispensed based on the molar ratio of the respective mixture. Each lipid or lipid mixture was dispensed in triplicate with wells H9-12 remaining empty to serve as a blank control (Table 2.3). The volatile solvents were evaporated for 12 hours under a constant stream of nitrogen gas to prevent oxidation. Afterwards, the lipid screen was sealed with an aluminium cover under a nitrogen atmosphere and stored at -20 °C in the dark until used. A total of 12 screens was produced to serve as a first test batch for the validation of the screen on three membrane proteins.

Table 2.3: Lipid screen design containing 31 different lipids or lipid mixtures. Adapted from Cecchetti and *et al.* (172).

Well	Lipid	Abbreviation	Amount	Lipid type
A1 A2 A3	1,2-Dioleoyl-*sn*-Glycero-3-Phosphoethanolamine	DOPE	0.3 mg	Phosphatidylethanolamine (PE)
A4 A5 A6	1-Palmitoyl-2-Oleoyl-*sn*-Glycero-3-Phosphoethanolamine	POPE	0.3 mg	Phosphatidylethanolamine (PE)
A7 A8 A9	1,2-Dipalmitoyl-*sn*-Glycero-3-Phosphoethanolamine	DPPE	0.3 mg	Phosphatidylethanolamine (PE)
A10 A11 A12	1-Palmitoyl-2-Oleoyl-*sn*-Glycero-3-Phosphocholine	POPC	0.3 mg	Phosphatidylcholine (PC)
B1 B2 B3	1,2-Dioleoyl-*sn*-Glycero-3-Phosphocholine	DOPC	0.3 mg	Phosphatidylcholine (PC)
B4 B5 B6	1,2-Dimyristoyl-*sn*-Glycero-3-Phosphocholine	DMPC	0.3 mg	Phosphatidylcholine (PC)
B7 B8 B9	1,2-Dihexadecanoyl-*sn*-Glycero-3-Phosphocholine	DPPC	0.3 mg	Phosphatidylcholine (PC)
B10 B11 B12	1-Myristoyl-2-Hydroxy-*sn*-Glycero-3-Phosphocholine	LMPC	0.3 mg	Phosphatidylcholine (PC)
C1 C2 C3	1-Palmitoyl-2-Oleoyl-*sn*-Glycero-3-Phosphoglycerol (sodium Salt)	POPG-Na	0.3 mg	Phosphoglyceride (PG)
C4 C5 C6	1,2-Dipalmitoyl-*sn*-Glycero-3-Phospho-1'-*rac*-Glycerol (sodium salt)	DPPG-Na	0.3 mg	Phosphoglyceride (PG)
C7 C8 C9	1-Palmitoyl-2-Oleoyl-*sn*-Glycero-3-Phospho-L-Serine (sodium salt)	POPS	0.3 mg	Phosphatidylserine (PS)
C10 C11 C12	1,2-Dioleoyl-*sn*-Glycero-3-Phospho-L-Serine (sodium salt)	DOPS-Na	0.3 mg	Phosphatidylserine (PS)
D1 D2 D3	1,2-Didecanoyl-*sn*-Glycero-3-Phosphate (sodium salt)	10:PA-Na	0.3 mg	Phosphatic acid (PA)
D4 D5 D6	1-Palmitoyl-2-Oleoyl-*sn*-Glycero-3-Phosphate (sodium salt)	16:0-18:1 PA (POPA)	0.3 mg	Phosphatic acid (PA)
D7 D8 D9	*E. coli* Polar Lipid Extract	EPL	0.3 mg	Mixture of PE (67%), PG (23.2%) and CA (9.8%)
D10 D11 D12	Cholesteryl Hemisuccinate Tris Salt	CHS	0.3 mg	Sterol
E1 E2 E3	Cholesterol	CHL	0.3 mg	Sterol
E4 E5 E6	1',3'-bis[1,2-dimyristoyl-*sn*-glycero-3-phospho]-glycerol (sodium salt)	14:0 CL	0.3 mg	Cardiolipin (CL)
E7		18:1 CL	0.3 mg	Cardiolipin (CL)

E8 E9	1',3'-bis[1,2-dioleoyl-*sn*-glycero-3-phospho]-glycerol (sodium salt)			
E10 E11 E12	Porcine Polar Brain Lipid Extract	PBL	0.3 mg	Mixture of PC (12.6%), PE (33.1%), PI (4.1%), PS (18.5%), PA (0.8%) and unknown (30.9%)
F1 F2 F3	Egg Sphingomyelin (chicken)	SM	0.3 mg	Mixture of sphingomyelins (SM) with 16:0 SM (86%), 18:0 SM (6%), 22:0 SM (3%), 24:1 SM (3%), unknown (2%)
F4 F5 F6	Monomyristolein	7.7MAG	0.3 mg	Monoacylglycerol
F7 F8 F9	Monoolein	9.9MAG	0.3 mg	Monoacylglycerol
F10 F11 F12	Lipid mix I	POPC:POPG:POPE (3:1:1)*	0.3 mg	Used for structure solution of several ion channels
G1 G2 G3	Lipid mix II	POPC:POPS (4:1)*	0.3 mg	Simple mimic of the membrane inner lipid leaflet
G4 G5 G6	Lipid mix III	POPG:POPE (3:1)*	0.3 mg	Simple mimic of Gram-positive bacterial membranes
G7 G8 G9	Lipid mix IV	POPG:POPE (1:3)*	0.3 mg	Simple mimic of Gram-negative bacterial membranes
G10 G11 G12	Lipid mix V	DMPC:CHL (2:1)*	0.3 mg	Simple mimic of eukaryotic plasma membranes
H1 H2 H3	Lipid mix VI	POPC:CHL (5:1)*	0.3 mg	Simple mimic of nerve cell membranes
H4 H5 H6	Lipid mix VII	CHL:POPC:SM (1.8:1:1)*	0.3 mg	Simple mimic of erythrocyte membranes
H7 H8 H9	Lipid mix VIII	POPC:POPE (1:1)*	0.3 mg	Simple mimic of mitochondrial membranes
H10 H11 H12	Blank	Blank	0 mg	Control

*molar ratio

2.3.2. Screen validation

The effect of membrane protein relipidation on the thermostability of three test proteins ($A_{2A}R$, UapA and *Tm*-PPase) was tested by differential scanning fluorimetry (DSF). Wild-type *Tm*-PPase was expressed and purified in octyl glucose neopentyl glycol (OGNG) as described in 2.6.1 and 2.6.2. Claudia Stohrer (University of Leeds, RAMP-ITN) supplied stabilised human $A_{2A}R$ construct (Δ317-412, 209-218 replaced with a thermostabilised apocytochrome b562 from *E. coli* (BRIL)). Cristina Cecchetti (Imperial College London) produced UapA-G411V$_{\Delta 1\text{-}11}$ for stability testing. The screen was thawed and centrifuged at 1000 *x*g for 3 minutes before 50 µL of 3% *w/v* detergent (DDM for $A_{2A}R$ and UapA, OGNG for *Tm*-PPase) were added for the solubilisation of 0.3 mg lipid,

followed by manual mixing with a multi-channel pipette. After 12 hours at 24 °C, shaking at 250 rpm, any undissolved lipids were collected at the bottom of the plate in a 3-minute centrifugation at 1,000 *x*g. 15 µL of the supernatant containing solubilised lipids were mixed with 15 µL of water and 15 µL of degassed 3x protein buffer ($A_{2A}R$: 30 mM HEPES pH 7.5, 2400 mM NaCl; UapA: 60 mM Tris-HCl pH 7.5, 450 mM NaCl, 1.8 mM xanthine; *Tm*-PPase: 150 mM MES-NaOH, pH 6.5, 11.5% *v/v* glycerol, 150 mM KCl, 15 mM $MgCl_2$), diluting the detergent and lipid concentration to 1% and 2 mg mL^{-1} (assuming that all lipid was solubilised), respectively. Purified protein at 1 mg mL^{-1} ($A_{2A}R$, *Tm*-PPase) or 0.5 mg mL^{-1} (UapA) was then mixed at a 1:1 ratio with the solubilised lipids to a final volume of 30 µL.

Stabilising effects of lipid addition were assessed in DSF experiments using the Prometheus NT.48 nanoDSF system (Nanotemper), which allows the simultaneous assessment of up to 48 different conditions with sample volume as little as 8-10 µL per measurement. Relipidated protein and detergent-solubilised protein samples with or without a known ligand added (*Tm*-PPase: 2 mM Na_4IDP, UapA: 1 mM xanthine, $A_{2A}R$: 50 µM ZM241385) were loaded into standard grade Prometheus NT.48 glass capillaries (Nanotemper). The instrument power was adjusted to give a maximal relative fluorescence unit (RFU) signal just below 20,000 in accordance with the manufacturer's instructions. A melting scan was set up ranging from 20 °C to 99 °C with a ramp rate of 1 °C min^{-1} using the PR.ThermControl (version 2.1.2) software. The fluorescence emission at 330 nm (F_{330}), 350 nm (F_{350}), the $F_{350:330}$ ratio and the light scattering signal were recorded over the course of the melting scan. For each protein, the effect of different lipids or ligand added was assessed in the same run by comparison to detergent-solubilised (apo) reference sample. The apparent melting temperatures (T_m) were obtained at the minimum/ maximum of the first derivative of the F_{330} signal, the F_{350} signal or the 350:330 ratio by the PR.ThermControl software and used as a means to assess protein thermostability. The experiment was repeated three times to calculate the standard error of the mean (SEM) that is stated in figures and throughout the text. The data was exported for plotting and statistical analysis (one-way ANOVA) in GraphPad Prism 9.0.

2.4. Thermostabilisation of *Cl*-PPase

2.4.1. Computational prediction of thermostabilising point-mutations

Thermostabilising point-mutations were predicted using all three modules (deep-sequence-based, structure-based and data-driven) of IMPROvER (173). The minimal requirement to run IMPROvER is the target protein sequence, here: the K^+-dependent Na^+,H^+-PPase from *Clostridium leptum (Cl*-PPase, UniProt: A7VNH8). A set of comparative homology models required for the structure-based prediction approach

were provided alongside the protein sequence for better performance. The comparative homology models were obtained from MODELLER (174) and I-TASSER (175) following the developer's instructions. Homology models generated by MODELLER were based on M-PPase structures available at the time (PDB: 4A01, 4AV3, 4AV6, 5GPJ, 5LZR, 5LZQ). I-TASSER was used to obtain an additional unrestricted model. IMPROvER compiled a ranked listed of putatively stabilising mutations of which the top 15% were selected for experimental testing. Any mutations overlapping with a list of residues critical for M-PPase integrity identified in a literature search were excluded (Table S5, Appendix).

2.4.2. *Cl*-PPase mutagenesis

Unless stated otherwise, enzymes and buffers for molecular cloning were purchased from New England Biolabs (NEB). The *Cl*-PPase construct used for stability screening was made synthetically (syn-*Cl*-PPase) and contains a N-terminal 8xHis-tag followed by superfolder green fluorescent protein (sfGFP), a TSAG linker, a Tobacco Etch Virus (TEV) protease cleavage sequence (ENLYFQ|G) and another TSAG linker upstream of the *Cl*-PPase encoding region. In addition, flanking homologous recombination domains (hrd) for insertion into the pDDGFP2 yeast expression plasmid (176) were added. Homologous recombination during transformation replaced an existing yGFP in the pDDGFP2 backbone and put the *Cl*-PPase gene under control of a galactose promoter. The experimental procedure followed the protocol described in 2.2.2 with minor modifications, namely the addition of 250 ng linearised pDDGFP2 plus 75 ng of syn-*Cl*-PPase instead of circular plasmid. Linearised plasmid was obtained by inverse PCR as described in 2.2.7 using back-to-back primers (#001-002) amplifying the entire template plasmid (Table S2, Appendix).

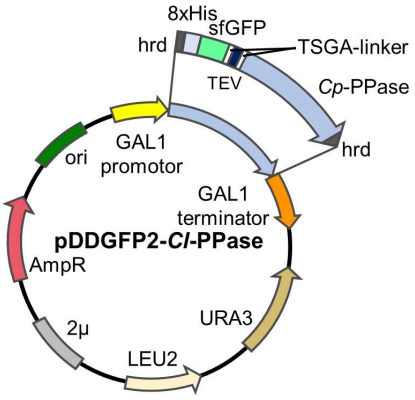


Figure 2.1: Schematic overview of pDDGFP2-*Cl*-PPase plasmid map.

Similarly, all point mutations were introduced into the sequence confirmed pDDGFP2-*Cl*-PPase plasmid (Figure 2.1) by inverse PCR using back-to-back mutagenesis primers (#005-199) containing the desired nucleotide changes (Table S2, Appendix). Site-directed mutagenesis based on IMPROvER prediction was done in a high-throughput format using 96-well PCR plates (Thermo Fisher Scientific) and subject to several iterations if no PCR product of correct size was obtained in the previous round. In the first round of Q5® site-directed mutagenesis, the average melting temperature of all mutagenesis primers (69 °C) was used for annealing. In subsequent rounds, primer pairs with similar melting temperatures were grouped to increase the success rate by

using more appropriate annealing temperatures in PCR. Template digestion and blunt-end ligation was done simultaneously by adding 2 µL of a 2:1:1:1 mix of DpnI, T4 polynucleotide kinase, T4 DNA ligase and 10x T4 DNA ligase buffer to 10 µL PCR product. The reaction took place at 20 °C for 1 hour and was stopped by heat inactivation at 80° C for 20 minutes. 2 µL of this mix were transformed into chemically competent OmniMax cells as described in 2.2.1. The presence of the *Cl*-PPase gene in picked transformants was confirmed by PCR using half of a bacterial colony resuspended in 10 µL nuclease free water as template. The primers (#003-004) used for colony PCR aligned on the hrd for the insertion of syn-*Cl*-PPase into pDDGFP2 (Table S2, Appendix). Plasmids were isolated by mini-prep (Macherey-Nagel) from positive colonies following the manufacturer's instructions and the correct introduction of point mutations was confirmed by DNA sequencing (Eurofins); first over the mutation site, then over the entire *Cl*-PPase encoding region.

2.4.3. Heterologous protein expression in *S. cerevisiae* and membrane extraction

8xHis-sfGFP-*Cl*-PPase was expressed following transformation of pDDGFP2-*Cl*-PPase into the *S. cerevisiae* strain FGY217 (genotype: *MATα, ura3-52, pepΔ4, lys2Δ201*) and uracil auxotrophic selection on synthetic complete dropout media minus leucine (SCD-Ura) plates (0.67% *w/v* yeast nitrogen base, 2% *w/v* D-glucose, 2% *w/v* agar, 0.1 mg mL^{-1} carbenicillin, 0.41 mg mL^{-1} L-threonine, 0.10 mg mL^{-1} L-phenylalanine, 0.06 mg mL^{-1} L-lysine, 0.04 mg mL^{-1} of L-arginine, 0.04 mg mL^{-1} L-histidine, 0.04 mg mL^{-1} L-leucine, 0.04 mg mL^{-1} L-methionine, 0.04 mg mL^{-1} L-tryptophan, 0.04 mg mL^{-1} L-tyrosine, 0.04 mg mL^{-1} adenine) as described in 2.2.2. Yeast starter cultures in SCD-Ura media were inoculated with a single colony and incubated at 30 °C for 16 hours, shaking at 250 rpm. The starter cultures were scaled up to 15 mL expression cultures for one-temperature GFP-based stability assays or 100 mL expression cultures for ten-temperature GFP-based stability assays. Expression cultures contained 0.1% *w/v* glucose instead of 2% *w/v* glucose, were inoculated at a starting OD_{600} of 0.15 and incubated at 30 °C until they reached an OD_{600} of 0.6. Protein expression was induced by addition of 2% *w/v* galactose and allowed to continue for 15 hours at 30 °C, shaking at 250 rpm. Afterwards, the cells were harvested by centrifugation (3,000 *x*g, 4 °C, 5 minutes) and pellets were resuspended in 750 µL lysis buffer (200 mM Tris-HCl pH 7.5, 40% *v/v* glycerol, 10 mM EDTA, 2 mM DTT, 0.2 mM phenylmethylsulfonylfluoride (PMSF), 2 µg/mL pepstatin A). Protein expression was quickly confirmed by recording the whole-cell GFP fluorescence at 520 nm in black 96-well clear-bottom microplates after excitation at 485 nm using a FLUOstar Optima microplate reader (BMG Labtech). The GFP signal was normalised to the OD_{600} at the timepoint of harvesting for analysis. Afterwards, the resuspended cells were transferred

into 2 mL tubes containing ~1.3 g of glass beads with a 0.5 mm diameter. Cell lysis was done by vortexing at maximum speed for 30 minutes at 4 °C using a Vortex-Genie 2 (Scientific Industries and a 24-tube adapter (Qiagen). Cell debris, intact cells and glass beads were removed from the lysate by centrifugation (8,000 *x*g, 4 °C, 1 minute). Subsequently, membranes were harvested from the supernatant in a high-speed centrifugation (36,000 *x*g, 4 °C, 1 hour)

2.4.4. GFP-based thermostability assay

FGY217 membrane pellets containing wild-type or variant GFP-tagged *Cl*-PPase were resuspended in 150 µL solubilisation buffer (50 mM MES-NaOH pH 6.5, 20% *v/v* glycerol, 37.5 mM KCl, 3.75 mM $MgCl_2$, 1 mM Na_4PP_i, 1.34% *w/v* DDM, 2 mM DTT, 0.2 mM PMSF, 0.15 µg/mL pepstatin A) for single-temperature GFP-based stability assays or 750 µL solubilisation buffer for ten-temperature GFP-based stability assays. Solubilisation took place over the course of 1.5 hours at 4 °C in 1.5 mL tubes that were continuously mixed by rolling. Residual insoluble material was removed by another high-speed centrifugation (36,000 xg, 4 °C, 1 hour). The fluorescence signal of the solubilised membranes was recorded using a QFX Fluorometer (DeNovix) to confirm the presence of GFP-tagged *Cl*-PPase. All samples were diluted to the lowest fluorescence reading obtained within a set of samples to allow direct comparison between them and to prevent over- or underexposure of different samples during gel imaging. Afterwards, samples were split into 60 µL aliquots of which one was kept at 4 °C to serve as a reference sample and the others were heated in a Thermomixer (Eppendorf) for 10 minutes. In one-temperature GFP-based stability assays heating was done at 50 °C only, whereas in ten-temperature stability assays a full melting curve was recorded by incubation at 20, 30, 40, 45, 50, 55, 60, 70, 80 and 90 °C. Samples were subsequently cooled on ice for 30 minutes prior to removing any precipitated protein by centrifugation (20,000 *x*g, 4 °C, 1 hour). 40 µL of supernatant were transferred into fresh tubes and samples were analysed by SDS-PAGE and Western-Blot using anti-M-PPase antibodies (1:10000 dilution, rabbit) as described in 2.2.4 and 2.2.5. *Cl*-PPase bands were visualised by in-gel fluorescence detection using the G:BOX with blue LEDs (Syngene). The full-length *Cl*-PPase band intensities were quantified using ImageJ to compare reference and temperature-challenged samples and plot melting curves. The T_m of wild-type and variant *Cl*-PPase was obtained at the inflection point of the plotted fluorescence signal (normalised to the on-ice sample) against the temperature challenge. Data were fitted to a four-parameter dose-response curve (variable slope) by non-linear least-squares fitting in GraphPad Prism 7.0. The SEM of T_m and ΔT_m were calculated where n≥3, with error propagation factored in for the latter, and stated in the figure and throughout the text. Statistical analysis was done by an ordinary one-way ANOVA with

a Dunnet follow-up test for multiple comparisons in GraphPad Prism 7.0 unless stated otherwise.

2.4.5. Activity measurement of *Cl*-PPase in membranes

FGY217 membranes containing GFP-tagged *Cl*-PPase were obtained from 100 mL SCD-Ura expression (2.4.3). The total protein concentration was determined in a Bradford assay as described in 2.2.6 and diluted to 2-4 mg mL^{-1} in 50 mM MES-NaOH pH 6.5, 20% *v/v* glycerol, 50 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 0.25 mM PMSF and 0.15 µg/mL pepstatin A. The hydrolytic activity of *Cl*-PPase in membranes was determined by quantifying the P_i release as reduced 12-molybdophosphoric acid following the protocol described in 2.2.8 with minor changes that are detailed below.

Reaction buffer, PP_i trigger solution and resuspended membranes were mixed at a 7:1:2 ratio to a final volume of 50 µL with 60-70 mM Tris-HCl pH 7.0, 3 mM $MgCl_2$, 115 mM KCl, 25 mM NaCl, 1 mM NaF, 5 µM Gramicidin D and 400 µM Na_2PP_i. All samples were prepared in triplicate and incubated at 30 °C for 10 minutes upon reaction initiation by the addition of PP_i trigger solution. In negative controls PP_i trigger solution was only added after the incubation at 30 °C when samples were transferred on ice. After assay development and absorbance measurement at 860 nm (2.2.8), the released P_i amount and specific protein activity were calculated. These calculations were based on the linear equation of the P_i calibration curve. *Cl*-PPase amounts in membranes were estimated based on its GFP signal and a GFP calibration curve, both recorded on a QFX Fluorometer (DeNovix). The SEM was obtained from three technical repeats.

2.5. Expression, purification and crystallisation of *Cl*-PPase

2.5.1. Heterologous protein expression in *E. coli*

Thermostabilised *Cl*-PPase used for structural studies was expressed in *E. coli*. Therefore, the most stabilising mutations suggested by IMPROvER were transferred into the bacterial expression plasmid pZH-*Cl*-PPase, which was made by former group member Dr. Nita Shah. Mutagenesis followed the protocol described in 2.4.2. To the N-terminal side of the *Cl*-PPase coding region is a maltose binding protein (MBP) signal sequence for membrane targeting, followed by an 8xHis-tag, MBP and a human rhinovirus 3C protease (HRV3C) cleavage site (LEVLFQ|GP). *Cl*-PPase expression is controlled by the T7/lac promoter system (Figure 2.2). Wildtype or variant pZH-*Cl*-PPase expression plasmid was transformed into BL21star (DE3) pRARE2 cells as described in 2.2.1. All media and agar mentioned below was supplemented with 100 µg mL^{-1}

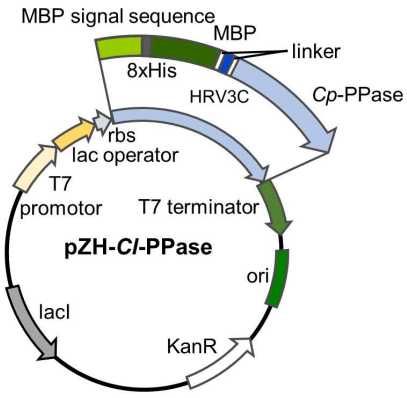


Figure 2.2: Schematic overview of the pZH *Cl*-PPase plasmid map.

kanamycin to enforce antibiotic selection pressure. Colonies grown on LB plates were used to inoculate 4x 250 mL of LB starter culture. After 18 hours at 37 °C and shaking at 250 rpm, 50 mL of starter culture were added to 500 mL freshly prepared TB media (starting OD_{600}: ~0.2). Up to 10 L expression culture were incubated at 37 °C while shaking at 250 rpm for ~6 hours until reaching an OD_{600} of 6-7. Prior to induction, all cultures were cooled down to 18 °C in an ice-water bath with occasional shaking. Protein expression was induced by addition of 0.5 mM IPTG and allowed to proceed for 1 hour at 18 °C, shaking at 250 rpm. Cells were harvested by centrifugation (4,500 *x*g, 4 °C, 15 minutes) and resuspended in PBS that was supplemented with 1 mM PMSF and 2 µg mL^{-1} pepstatin A to a volume not exceeding 200 mL. The cell suspension was equally split between 4x 50 mL Falcon tubes. Washed cells were collected at 4,500 *x*g and 4 °C for 15 minutes, flash frozen in liquid nitrogen and stored at -80°C until needed.

To compare protein expression between wildtype and variant *Cl*-PPase the equivalent of 1 mL at an OD_{600} of 0.5 was taken from 0.5 L expression cultures incubated in parallel. Cells were harvested by centrifugation (4,500 *x*g, 4 °C, 5 minutes) and resuspended in 50 µL PBS. For cell lysis, each resuspended pellet was incubated at 60 °C for 5 minutes. 1x reducing protein loading buffer was added to each tube and the cell debris and DNA was removed by centrifugation (10,000 *x*g, 4°C, 5 minutes). 5, 10 and 20 µL of supernatant from each sample were loaded onto a Mini-Protean® TGX (4--20%) precast gel (Bio-Rad) for Western-Blot analysis (2.2.5) with α-His-HRP antibody (1:5000 dilution). For expression level comparison the Western-Blot signal was normalised to the total protein amount on the membrane quantified by Ponceau S staining (0.1% w/v Ponceau S, 1% v/v acetic acid) and densiometric analysis in ImageJ.

2.5.2. Membrane extraction and protein purification

Cells were resuspended in lysis buffer (200 mM Tris-HCl pH 7.5, 40% *v/v* glycerol, 10 mM EDTA, 2 mM DTT, 1 mM PMSF, 2 µg/mL pepstatin A,) using 2 mL buffer per gram of cell pellet and homogenised on ice. Cell lysis was done by high pressure homogenisation (3x at 20 kpsi) in a Avestin Emulsiflex C3 (ATA Scientific) cell disruptor. The lysate was supplemented with 330 mM KCl and the glycerol concentration was diluted to 15% *v/v* using lysis buffer lacking glycerol. The cell debris and intact cells were then removed by centrifugation (4,500 *x*g, 4 °C, 15 minutes). Afterwards, membranes were collected from the supernatant by ultracentrifugation (100,000 *x*g, 4 °C, 1 hour) and resuspended in 50 mM MES-NaOH pH 6.5, 20% *v/v* glycerol, 50 mM KCl, 5 mM $MgCl_2$, 1.33 mM DTT, 1 mM PMSF and 2 µg/mL pepstatin A. The total membrane protein concentration was determined using the Bradford assay as described in 2.2.6 and diluted to 15 mg mL^{-1} in resuspension buffer. The membrane suspension was mixed with 4x solubilisation buffer (50 mM MES-NaOH pH 6.5 20% *v/v* glycerol, 5.34% DDM, 1 mM

Na_4PP_i) in a 1:3 ratio. Solubilisation took place over the course of 12 hours at 4 °C in 50 mL Falcon tubes that were continuously mixed by rolling. Residual insoluble material was removed by ultracentrifugation (100,000 *x*g, 4 °C, 1 hour). *Cl*-PPase was purified by immobilised metal affinity chromatography (IMAC) with 1 mL cobalt nitrilotriacetic-acid (NTA) added for each 40 mL of solubilised membranes. The resin was mixed for 2.5 hours at 4 °C in 50 mL Falcon tubes by rolling and transferred into Econo-Pac gravity flow columns (Bio-Rad). Afterwards, it was washed twice with 12 column volumes (CV) HRV3C cleavage buffer (50 mM MES pH 6.5, 3.5% *v/v* glycerol, 50 mM KCl, 5 mM $MgCl_2$, 0.05% *w/v* DDM, 1 mM DTT) and then resuspended in 0.8 CV cleavage buffer. The resin resuspension was then transferred into 15 mL Falcon tubes and mixed with HRV3C at a 1:1000 ratio based on the total membrane protein amount determined in the Bradford assay earlier. HRV3C cleavage was done for either 4 hours or 12 hours at 4 °C while continuously mixing by rolling. The resin was transferred back into a gravity flow column, the flow-through was collected and residual protein was washed off by rinsing with 0.5-1 CV cleavage buffer. The protein concentration was determined by absorbance reading at 280 nm (Denovix DS-11 Spectrometer). If required, the protein was concentrated in Vivaspin ultrafiltration units (Sartorius) with polyether sulfone (PES) membranes and a molecular weight cut-off (MWCO) of 50 kDa following the manufacturer's instructions. Throughout the purification small aliquots were collected and analysed by SDS-PAGE and Western-Blot (α-His-HRP antibody) as described in 2.2.4 and 2.2.5.

2.5.3. Detergent screening and purification optimisation

Unless stated otherwise, *Cl*-PPase expression and purification followed the protocol described in 2.5.1 and 2.5.2. For detergent screening, wild-type or variant (F20Y, G130A) *Cl*-PPase was expressed in *E. coli*, solubilised in DDM and then exchanged into a set of mostly shorter-chain detergents that are favourable for vapour diffusion crystallisation; namely DDM, OGNG, n-undecyl-β-D-thiomaltopyranoside (UTM), n-decyl-α-D-maltopyranoside (DM), Cymal-6 (C6), n-nonyl-β-D-maltopyranoside (NM), n-nonyl-β-D-glucopyranoside (NG), n-octyl-β-D-maltopyranoside (OM), n-octyl-β-D-glucopyranoside (OG) and lauryl dimethylamine-N-oxide (LDAO). After solubilisation, protein-bound resin was split into equal parts and transferred into different gravity flow columns. The HRV3C cleavage buffer used for washing, proteolytic cleavage and protein elution was prepared with DDM, UTM, DM, C6, NM, OGNG, NG, OM, OG or LDAO at 3x CMC. Resin, which was resuspended in an identical volume of cleavage buffer, and eluted protein samples were analysed by SDS-PAGE and Western-Blot (α-His-HRP antibody, 1:5000 dilution) as described in 2.2.4 and 2.2.5. DDM served as an internal reference sample to correct for yield differences between individual purification batches and relative protein yields to

the purification in DDM were obtained by densiometric analysis of *Cl*-PPase bands in ImageJ.

The large-scale *Cl*-PPase purification protocol developed by Dr. Nita Shah was optimised for the incubation period of detergent solubilisation, protein to resin binding and HRV3C cleavage, as well as for the ratio of HRV3C to protein. All other purification steps were carried out as detailed in 2.5.2. Samples were split into equal parts and treated the same except for the optimisation parameter. The solubilisation efficiency at 4 °C was assessed over the time course of 1, 2, 4, 6 and 15 hours. The protein to resin binding at 4 °C was monitored over an incubation period of 0.5, 1, 2 and 4 hours. HRV3C cleavage was tested at a protease to protein ratio of 1:500, 1:1000, 1:2000, 1:4000 and 1:8000 and allowed to proceed for 1, 2, 4 or 15 hours at 4°C. All samples collected throughout the purification were analysed by SDS-PAGE and Western-Blot (α-His-HRP antibody) as described in 2.2.4 and 2.2.5. The efficiency of solubilisation, protein to resin binding and HRV3C cleavage was evaluated by densiometric analysis of *Cl*-PPase bands in ImageJ.

2.5.4. Activity measurements of wild-type and variant *Cl*-PPase in different detergents

Hydrolytic activity measurements of wild-type and variant *Cl*-PPase purified in a range of different detergents (2.5.1-2.5.3) were done by quantifying the P_i release as reduced 12-molybdophosphoric acid following the protocol described in 2.2.8 with minor modifications. Initially, wild-type *Cl*-PPase purified in DDM was used to optimise the incubation period at 30 °C and protein concentration to ensure P_i release is linear with time and falls in the range of the calibration curve. Relipidated protein at 0.4-3.2 mg mL^{-1} was 200x diluted in reaction buffer and mixed with PP_i trigger solution at a 4:1 ratio to a final volume of 50 µL with 50-60 mM Tris-HCl pH 7, 5 mM $MgCl_2$, 100 mM KCl, 10 mM NaCl and 400 µM Na_2PP_i. Negative controls only contained protein buffer without any protein. The enzymatic reaction was allowed to proceed for 10-30 minutes at 30 °C. The effect of different point mutations or detergents on protein activity was assessed using optimised assay parameters that are detailed below. Relipidated protein at 1.2 mg mL^{-1} was 200x diluted in reaction buffer and mixed with PP_i trigger solution at a 4:1 ratio to a final volume of 50 µL with 50-60 mM Tris-HCl pH 7, 5 mM $MgCl_2$, 100 mM KCl, 10 mM NaCl and 400 µM Na_2PP_i. The enzymatic reaction was allowed to proceed for 20 minutes at 30 °C before samples were placed on ice for stopping. The sample treatment was adjusted for protein purified in a range of different detergents. For protein relipidation, the detergent concentration was reduced from 600x CMC DDM (standard) to 1x CMC detergent, whereas the reaction buffer was supplemented with 3x CMC detergent to prevent protein from precipitating upon its dilution in reaction buffer. The chosen detergent matched the detergent that *Cl*-PPase was purified in. After assay development

and absorbance measurement at 860 nm (2.2.8), the released P_i amount and specific protein activity were calculated with the SEM obtained from three technical repeats.

2.5.5. Lipidic cubic phase crystallisation of *Cl*-PPase

F20Y and G130A *Cl*-PPase variants were purified in DDM from 5 L of expression culture as described in 2.5.1 and 2.5.2. 20 µL of sample at 20 mg mL^{-1} (F20Y) or 25 mg mL^{-1} (G130A) were supplemented with 4 mM Na_4IDP and incubated on ice for 1 hour. Any precipitation that formed during the incubation was removed by centrifugation (20,000 *x*g, 4 °C, 15 minutes). LCP was prepared following the standard protocol (62) with 2x 100 µL Hamilton syringes connected *via* a coupler and 9.7MAG as host lipid. The equivalent mass of 20 µL 9.7MAG was molten at 40 °C and transferred into a warm Hamilton syringe. The other Hamilton syringe was kept cool and loaded with 20 µL of purified protein. Both syringes were connected using a coupler and LCP was formed by steady mixing at about 1-2 strokes per second (1:1 ratio of 9.7MAG to protein). 50 nL LCP were dispensed with a NT8® crystallisation robot (Formulatrix) onto glass sandwich plates (Jena Biosciences) and overlaid by 800 nL of mother liquor. The following commercial sparse-matrix screens from Molecular Dimensions were diluted to 70% *v/v* for better compatibility with LCP, unless stated otherwise, and used in screening: MemGold1, MemGold 2, MemTrans, MemSys, MemStart and MemMeso (undiluted). Additionally, a custom screen designed to optimise a crystallisation condition (100 mM trisodium citrate pH 5.0, 50 mM NH_4Cl, 30% *v/v* PEG500 DME) in which *Pa*-PPase LCP crystals form (4.2.2) was set up alongside the commercial sparse-matrix screens. Varying PEG500 DME concentrations (24-34% *v/v*) were screened against different salt types (NH_4Cl, $(NH_4)_2SO_4$, NaCl+$LiSO_4$ at 1:1 ratio), salt concentrations (25-300 mM) and pH (100 mM trisodium citrate pH 5.5-7.0, 0.5 increments). All plates were sealed with glass covers and stored at 20 °C in a temperature controlled automated imaging system (RockImager® 1000, Formulatrix) for regular monitoring of crystal formation.

2.6. Expression, purification and crystallisation of *Pa*-PPase and *Tm*-PPase

2.6.1. Heterologous expression of thermostable M-PPases in *S. cerevisiae*

Yeast expression plasmids carrying N-terminally 6xHis-tagged *Pa*-PPase or *Tm*-PPase under control of the constitutively active PMA1 promoter were freshly transformed into the *S. cerevisiae* strain BJ1991 (genotype: *MATα prb1-1122 pep4-3 leu2 trp1 ura3-52 gal2*) as described in 2.2.2 and cultivated at 30 °C for 12 hours in 250 mL 1x SCD-Leu media (0.67% *w/v* yeast nitrogen base, 2% *w/v* D-glucose, 0.1 mg mL^{-1} carbenicillin, 0.06 mg mL^{-1} L-histidine, 0.1 mg mL^{-1} L-tryptophan, 0.035 mg $mL^{-1,}$ adenine, 0.022 mg mL^{-1} uracil). The constructs were made by former group members Dr. Craig Wilkinson and Dr. Juho Kellosalo. 250 mL starter cultures were then mixed with 750 mL

1.5x SCD-Leu media in 2.5 L baffled flasks. Protein expression took place over 10 hours at 30 °C and 200 rpm shaking. Cells were harvested from 10 L expression batches by centrifugation (4,000 *x*g, 4 °C, 15 minutes) and resuspended in water to a volume not exceeding 200 mL. The cell suspension was equally split between 4x 50 mL Falcon tubes. Washed cells were collected at 4,500 *x*g and 4 °C for 15 minutes, flash frozen in liquid nitrogen and stored at -80°C until used.

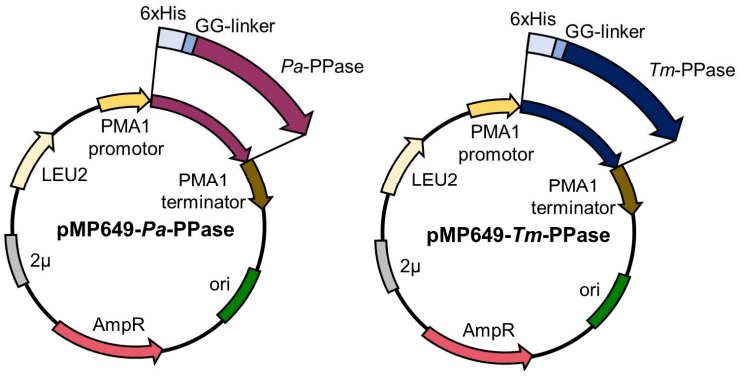


Figure 2.3: Schematic overview of the pMP649 plasmid maps for thermostable M-PPase expression in yeast. *Pa*-PPase (left panel) or *Tm*-PPase (right panel).

2.6.2. Membrane extraction and hot-solve purification

Cell pellets were thawed and resuspended in 1 mL lysis buffer (200 mM Tris-HCl pH 7.5, 40% *v/v* glycerol, 10 mM EDTA, 2 mM DTT, 1 mM PMSF, 2 µg/mL pepstatin A) per gram cell mass. Lysis was done at 4 °C using a bead-beater (Biospec products) with 0.5 mm glass beads by 24x 1-minute activations interspersed by 30-second cooling periods on ice. The lysate was supplemented with 330 mM KCl and the glycerol concentration was diluted to 15% *v/v* using lysis buffer that had no glycerol added. Afterwards, cell debris and intact cells were removed by centrifugation (4,500 *x*g, 4 °C, 15 minutes). Then, membranes were collected from the supernatant by ultracentrifugation (100,000 *x*g, 4 °C, 1 hour) and resuspended in 50 mM MES pH 6.5, 20% *v/v* glycerol, 50 mM KCl, 5 mM $MgCl_2$, 2 mM DTT, 1 mM PMSF and 2 µg/mL pepstatin A. The total membrane protein concentration was determined by the Bradford assay as described in 2.2.6 and diluted to 7 mg mL^{-1} in resuspension buffer. Protein solubilisation and purification followed the “hot-solve” method developed for the isolation of recombinant thermophilic M-PPases (177). The membrane suspension was mixed with 4x solubilisation buffer (50 mM MES-NaOH pH 6.5 20% *v/v* glycerol, 5.34% DDM, 1 mM Na_4PP_i) at a 3:1 ratio. Solubilisation took place at 75 °C for 1.5 hours. Subsequently, insoluble matter was removed by centrifugation (4,500 xg, 20 °C, 15 minutes). *Tm*-PPase was purified by IMAC using 0.5 mL nickel-NTA for each 40 mL of solubilised membranes. Resin was transferred into Econo-Pac gravity flow columns (Bio-Rad) and was washed with 2 CV 50 mM MES-NaOH pH 6.5, 20% *v/v* glycerol, 50 mM KCl, 5 mM $MgCl_2$, 20 mM imidazole, 1 mM DTT and detergent (specified below). After washing, the protein was eluted in 2 CV 50 mM MES-NaOH, pH 6.5, 3.5% *v/v* glycerol, 5 mM $MgCl_2$, 400 mM imidazole, 1 mM DTT and detergent. Purified *Tm*-PPase was exchanged into elution

buffer lacking imidazole (exchange buffer) using a PD10 desalting column (Cytiva) according to the manufacturer's instructions. *Pa*-PPase purification followed the same protocol without KCl added to the elution and exchange buffer. This is because *Pa*-PPase is K^+-independent. The detergent used throughout IMAC and buffer exchange varied depending on downstream experiments. For vapour diffusion crystallisation of *Pa*-PPase, buffers contained 0.5% *w/v* DM. *Tm*-PPase was purified in 0.5% *w/v* OGNG instead. For activity measurements and lipidic cubic phase crystallisation of *Tm*-PPase and *Pa*-PPase, 0.05% *w/v* DDM was used unless stated otherwise. The protein concentration was determined by absorbance reading at 280 nm (Denovix DS-11 Spectrometer). If required, the protein was concentrated in 100 kDa MWCO Vivaspin ultrafiltration units with PES membranes (Sartorius) following the manufacturer's instructions. Throughout the purification small aliquots were collected and analysed by SDS-PAGE and Western-Blot (α-His-HRP antibody, 1:5000 dilution) as described in 2.2.4 and 2.2.5.

2.6.3. Size exclusion chromatography

Protein obtained by the "hot-solve" purification method as described in 2.6.2 was concentrated to 500 µL using 100 kDa MWCO Vivaspin ultrafiltration units (Sartorius) and any precipitation was removed by subsequent centrifugation (20,000 *x*g, 4 °C, 15 minutes). Samples were injected onto a Superose® 6 Increase 10/300 GL column (Cytvia) connected to an NGC Quest 10 Plus System (Bio-Rad) for size separation at room temperature. The column was pre-equilibrated in the respective protein buffer and run at 0.5 mL/min with 0.5 mL fractions being collected in a deep well plate.

2.6.4. Activity measurements of purified *Pa*-PPase and *Tm*-PPase

Most M-PPase purification batches obtained in 2.5.1 and 2.5.2 were tested for hydrolytic activity. This was done by quantifying the P_i release as reduced 12-molybdophosphoric acid following the protocol described in 2.2.8. Relipidated protein at 0.4 mg mL^{-1} was 200x diluted in reaction buffer and mixed with PP_i trigger solution at a 4:1 ratio to a final volume of 50 µL with 50-60 mM Tris-HCl pH 7.0, 5 mM $MgCl_2$, 100 mM KCl, 10 mM NaCl and 400 µM Na_2PP_i. Protein buffer without any protein was used as negative control. All samples were prepared in triplicate. The enzymatic reaction was allowed to proceed for 5 minutes at 70 °C (*Tm*-PPase) or 80 °C (*Pa*-PPase). The assay was developed and the absorbance at 860 nm read (2.2.8). Afterwards, the released P_i amount and specific protein activity were calculated based on the linear equation of the standard curve.

2.6.5. Vapour diffusion and lipidic cubic phase crystallisation of *Pa*-PPase

Vapour diffusion and LCP crystallisation trials were set up with wild-type *Pa*-PPase purified in DDM from 10 L expression batches as described in 2.6.1 and 2.6.2. Protein in its apo form and bound to non-hydrolysable inhibitor (2 mM Na_4IDP) was used in

crystallisation trials. Any precipitation that formed during incubation with Na_4IDP (4 °C, 1 hour) was removed by centrifugation (20,000 *x*g, 4 °C, 15 minutes) prior to setting up crystallisation plates.

Vapour diffusion crystallisation focused on optimising conditions that previously yielded *Pa*-PPase crystals (100 mM MES-NaOH pH 6.5, 30% *v/v* PEG400, 50 mM $LiSO_4$, 50 mM NaCl) diffracting to up to 3.8 Å (178) and several *Tm*-PPase crystals diffracting between 2.6-4.0 Å (127,129). Varying PEG400 concentrations (29-39% *v/v*) were screened against different salt types (NaCl, KCl, $CaCl_2$, NH_4Cl, NaCl+$LiSO_4$ at a 1:1 ratio), salt concentrations (50-200 mM) and pH (100 mM MES-NaOH/ Tris-HCl at pH 5.5-8.5, 0.2 increments) in 96-well MRC 3-drop plates. 200 nL drops of *Pa*-PPase at 5-10 mg mL^{-1} were mixed with mother liquor at a 1:1 ratio and set up using a NT8® crystallisation robot (Formulatrix). The reservoir volume was 25 µL. All plates were sealed and stored at 20 °C in a temperature controlled automated imaging system (RockImager® 1000, Formulatrix) for regular monitoring of crystal formation.

LCP crystallisation plates were set up with protein at 18-25 mg mL^{-1} in 9.9MAG and 9.7MAG. To minimise protein loss during concentration, the Cubicon method for stepwise protein accumulation in LCP was followed (67). Host lipid was melted at 40 °C and transferred into a warm Hamilton syringe. Another Hamilton syringe was kept cool and loaded with protein using a 1:1.5-3.0 ratio of lipid to protein (excess protein solution). Both syringes were connected using a coupler and LCP was formed by steady mixing at about 1-2 strokes per second. The usage of excess protein solution led to the formation of a dispersion of hydrated LCP (turbid) in aqueous solution by thorough mixing to reach equilibrium in this two-phase system. During mixing, the additional protein transitioned into LCP, which was then physically separated from the protein-depleted aqueous phase by slowly pushing it from one syringe into the other through the narrow coupler that acted as a filter for the more viscous LCP. The protein-depleted aqueous phase was exchanged with new protein sample and the was process repeated several times until a concentration of 10-15 mg mL^{-1} was reached in LCP. Optically clear LCP used in crystallisation trials was formed by adding a certain amount of host lipid back once the last round of concentration was finished and protein-depleted aqueous phase removed. The required amount of lipid was calculated based on the starting amounts of lipid and protein solution, the volume of the hydrated LCP after concentration and the specific lipid to water ratio for LCP formation of the lipid used (9.9MAG: 3:2, 9.7MAG: 1:1). The protein-depleted aqueous phases of each round were analysed by SDS-PAGE as described in 2.2.4 to monitor the accumulation of protein in LCP throughout the process.

Either 50 or 100 nL LCP were then dispensed with an NT8® crystallisation robot (Formulatrix) onto glass sandwich plates (Jena Biosciences) and overlaid with 800 nL or

400 nL of mother liquor, respectively. The following commercial sparse-matrix screens from Molecular Dimensions were diluted to 70% *v/v* to be better compatible with LCP, unless stated otherwise, and used in initial screens: MemGold1, MemGold2, MemTrans, MemSys, MemStart and MemMeso (undiluted). Further optimisation was carried out based on a screening condition (70 mM trisodium citrate pH 5.0. 70 mM NaCl and 21% *v/v* PEG500 DME) showing microcrystal formation. This included varying PEG500 DME or PEG400 concentration (20-34% *v/v)* against different salt types (NaCl, KCl, NH_4Cl, $(NH_4)_2SO_4$, NaCl+$LiSO_4$ at 1:1 ratio, NaCl+KCl at 1:1 ratio), salt concentrations (25-300 mM) and pH (100 mM trisodium citrate, pH 4.5-7.5, 0.5 increments). Additionally, the effect of additives was tested using a 1:10 dilution of the Hampton Additive Screen™ HT in 100 mM trisodium citrate pH 5.0,100 mM NH_4Cl and 26% *v/v* PEG500 DME. Besides glass sandwich plates, a range of *in situ* plates were used to set up crystallisation trials as well, namely the MiTeGen In situ-1 ™ plates, MiTeGen IMISX™ plates and prototype Swiss CI LCP crystallisation films. All plates were sealed and stored at 20 °C in a temperature controlled automated imaging system (RockImager® 1000, Formulatrix) for regular monitoring of crystal formation.

2.6.6. Vapour diffusion and in batch crystallisation of *Tm*-PPase for time-resolved structural studies

Tm-PPase was purified from 10 L expression batches in OGNG and concentrated to 10 mg mL^{-1} as described in 2.6.1 and 2.6.2 with minor modifications. In the exchange buffer, MES was pH adjusted with tetramethylammonium hydroxide (TMA-OH) instead of NaOH to avoid Na^+ contamination. Na^+ will be used to trigger the reaction initiation later. Initial vapour diffusion crystallisation trials of *Tm*-PPase for time-resolved structural studies were based on the crystallisation condition (36% *v/v* PEG 400, 100 mM Tris-HCl pH 8.5, 100 mM $MgCl_2$, 100 mM NaCl, 2 mM DTT) that yielded the *Tm*-PPase:Mg_5IDP structure (129). 1-2 µL crystallisation drops were set up at a 1:1 ratio of purified protein to mother liquor in 24-well hanging drop plates with a reservoir volume of 500 µL. Initially, crystals of this condition were reproduced using protein supplemented with 4 mM Na_4IDP, which was then replaced by K_4PP_i in subsequent crystallisation trials. Any precipitation that formed during incubation with ligand (4 °C, 1 hour) was removed by centrifugation (12,000 *x*g, 4 °C, 15 minutes) prior to setting up plates. Additionally, NaCl was substituted with KCl or CsCl in the crystallisation condition to lock the protein in the absence of Na^+ but in the presence of its substrate PP_i in the resting state without using inhibitors. For crystal optimisation, varying PEG400 concentrations (24-44% *v/v*) were screened against different salt concentrations ($MgCl_2$: 2-20 mM, KCl: 50-220 mM), pH (pH 8-9, 0.5 increments) or pre-crystallisation sample treatment (co-crystallisation with 0.4-4 mM K_4PP_i or addition of 2 mM L-α-lecithin from soybean). The effect of different PEGs (PEG200, PEG300, PEG400, PEG 500DME) on crystal growth was tested as well.

All crystallisation plates were stored at 20 °C and manually inspected daily to monitor crystal formation. Crystals used for time-resolved structural studies were finally obtained in 24-26% *v/v* PEG 400, 50-60 mM Tris-HCl pH 8.5, 2-3 mM $MgCl_2$, 175 mM KCl and 2 mM DTT (2.8.2).

Batch crystallisation was based on two vapour diffusion conditions suitable for time-resolved structural studies of *Tm*-PPase: the very first condition that gave suitable crystals (32-44% *v/v* PEG400, 100 mM Tris-HCl pH 8.5, 100 mM $MgCl_2$, 100 mM KCl, 2 mM DTT) and the final condition after optimisation (26% *v/v* PEG400, 50 mM Tris-HCl pH 8, 175 mM KCl, 2 mM $MgCl_2$). In batch crystallisation, protein and crystallisation solution are typically mixed in a 1:1-1:3 ratio. Therefore, the precipitant concentration needs to be increased to match that of promising vapour diffusion conditions after mixing (dilution) so as to hit the nucleation zone from which crystallisation starts (179).

Following a recent publication (179), only the PEG400 and K_4PP_i concentration were increased in batch experiments, which were based on the initial vapour diffusion conditions yielding *Tm*-PPase crystals for time-resolved structural studies, to account for the dilution upon mixing. *Tm*-PPase at 10 mg mL^{-1} was supplemented with 6 mM K_4PP_i and mixed with 64% *v/v* PEG400, 50 mM Tris-HCl pH 8.0, 175 mM KCl and 2 mM $MgCl_2$ crystallisation solution at a 1:2 ratio to a final volume of 30-45 µL. The samples were stored at 20 °C in sealed Eppendorf tubes and regularly monitored for crystal formation. During optimisation, the PEG400 concentration was varied between 45-64% *v/v*. If heavy precipitation formed within the first 1-2 hours, 5-15 µL of crystallisation buffer lacking precipitant were added to steer the condition towards the nucleation zone.

In contrast, batch crystallisation based on the optimised vapour diffusion condition was set up to exactly match the latter after mixing protein with batch crystallisation solution at a 1:2 ratio. Therefore, the concentrations of all components in the batch crystallisation solution, not just PEG400, were increased accordingly. This was done to preserve the already optimised crystallisation conditions from vapour-diffusion experiments. *Tm*-PPase at 10 mg mL^{-1} was supplemented with 0.6 mM K_4PP_i and mixed with 39% *v/v* PEG400, 75 mM Tris-HCl pH 8.0, 262 mM KCl and 3 mM $MgCl_2$ at a 1:2 ratio to a final volume of 30 µL. The samples were either incubated for 2 minutes at 4 °C to promote nuclei formation or immediately stored at 20 °C in sealed Eppendorf tubes and regularly monitored for crystal formation.

2.7. Structural studies and functional characterisation of *Pa*-PPase

2.7.1. Structure solution and refinement

Pa-PPase crystals grown by vapour diffusion or in LCP (2.6.5) were fished using harvesting loops of appropriate type and size, cryo-cooled in liquid nitrogen and shot at

beamlines i04 and i24 at the Diamond Light Source (DLS). Experiments leading to the data collection of the 3.8 Å *Pa*-PPase dataset were carried out by former group member Dr. Craig Wilkinson (178). In brief, protein expression, purification and crystallisation followed the protocols described in 2.6.1, 2.6.2 and 2.6.5. *Pa*-PPase crystals were sent to several beamlines including i04 and i24 at the DLS and ID23-1 and MASSIF-1 at the European Synchrotron Radiation Facility (ESRF) for data collection. Collected datasets were processed in XDS (180) and the structure was solved by molecular replacement in Phaser (181) using a homology search model from I-TASSER (175), which was based on the TmPPase:Mg_5IDP structure (PDB: 5LZQ) with loop regions removed. The crystals were extremely radiation sensitive. Therefore, the first few hundred images of eight datasets (3.8-4.4 Å) with positive density for Mg_5IDP in the active site, less than 2% deviation in unit cell parameters and identical space group ($P2_1$) were combined in XDS using XSCALE (180) to reduce the effect of radiation damage-induced resolution loss.

I continued the structure solution, refinement and analysis of the 3.8Å combined dataset as detailed below. The combined dataset was submitted to the StarAniso webserver to correct for anisotropy (182) prior to MR as described above. Several rounds of refinement using phenix.refine (183) and manual modelling in Coot (184) were carried out. After an initial round of rigid-body refinement with grouped B-factors, tight restraints were applied to maintain realistic geometry (torsion angle non-crystallographic symmetry (NCS), secondary structure, and reference structure (PDB: 4A01) restraints). Additionally, X-ray/B-factor and X-ray/stereochemistry weights were optimised by phenix. All but the torsion angle NCS restraints were released in the last rounds of refinement to prevent overfitting of the model and Translation/ Libration/ Screw (TLS) was enabled.

2.7.2. Structural comparison and analysis

Geneious R11 was used to search the UniProtKB/Swiss-Prot database with blastp (185) for similar sequences to *Pa*-PPase and the results were aligned using the Geneious global alignment tool with free end gaps to determine residue conservation and sequence identity. Structure alignments were done and the rmsd/C_α was calculated in PyMol 2.2.3 (186). The solvent accessible surface areas and volumes were determined using HOLLOW with a 1.4-1.5 Å interior probe size (187). Inter-atom difference distance matrices of Cα atoms were generated by the Bio3D R-package for structural bioinformatics (188). Hydrogen bonding patterns were analysed in HBplus using default settings (189). The local (residue by residue) helix curvature analysis was done considering blocks of 4 residues using the Bendix plugin of the Visual Molecular Dynamics suite (190), whereas the global (helix by helix) curvature analysis was done using the HELANAL-Plus webserver(191).

2.7.3. Activity measurements assessing substrate inhibition and K^+-dependence

To record substrate inhibition curves and assess M-PPase dependence on K^+ as a co-factor, wild-type *Pa*-PPase was expressed and purified as described in 2.6.1 and 2.6.2 with minor modifications. MES was pH adjusted with TMA-OH in all purification buffers instead of NaOH to avoid Na^+ contamination due to its ability to substitute for K^+. Additionally, all purification buffers also lacked KCl. In K^+-(in)dependence assays, a PP_i trigger solution at 2 mM was prepared by titrating pyrophosphoric acid with TMA-OH to pH 7.0 instead of using Na_4PP_i. The P_i release from reactions with varying substrate concentration (0-1400 µM Mg_2PP_i) or reaction conditions (± K^+/Na^+) was quantified as reduced 12-molybdophosphoric acid following the protocol described in 2.2.8. Relipidated protein at 0.4-0.8 mg mL^{-1} was 200x diluted in reaction buffer and mixed with PP_i trigger solution at a 4:1 ratio to a final volume of 50 µL. Negative controls contained protein buffer without any protein. Initially, the incubation period (0.5-10 minutes) at 80 °C for wild-type *Pa*-PPase and *Tm*-PPase was optimised to ensure P_i release is linear with time and falls in the range of the P_i calibration curve. In subsequent assays at 80 °C, the incubation time was adjusted to 8 minutes. The final reaction condition used for optimisation was 60 mM Tris-HCl pH 7.0, 10 mM Na^+, 100 mM KCl, 5 mM free Mg^{2+} and 75/200 µM Mg_2PP_i (*Pa*-PPase/*Tm*-PPase). In substrate inhibition assays, the substrate concentration varied between 0 and 1400 µM Mg_2PP_i. In K^+-dependence assays, the reaction buffer either lacked KCl, NaCl or both, whereas the substrate concentration was kept constant at 75/200 µM Mg_2PP_i (*Pa*-PPase/*Tm*-PPase). The concentrations required to obtain 5 mM free Mg^{2+} and the desired Mg_2PP_i concentration were approximated as described by Baykov and co-workers (171). After assay development and absorbance measurements at 860 nm, the released P_i amount and specific protein activity were calculated with SEM obtained from three technical repeats.

2.8. Time-resolved structural studies of *Tm*-PPase

2.8.1. Activity measurements in crystallisation conditions

The hydrolytic activity of purified *Tm*-PPase used for time-resolved structural studies (2.6.6) was assessed in a range of different reaction buffers that mirrored crystallisation conditions to estimate time scales of substrate turnover by crystalline sample. The hydrolytic activity was determined by quantifying the P_i release as reduced 12-molybdophosphoric acid following the protocol described in 2.2.8. As a reference point, *Tm*-PPase activity was assessed in routine reaction conditions (2.2.8). Relipidated protein at 0.4 mg mL^{-1} was 200x diluted in reaction buffer and mixed with PP_i trigger solution at a 4:1 ratio to a final volume of 50 µL with 60 mM Tris-HCl pH 8.0, 20 mM NaCl, 100 mM KCl, 3 mM $MgCl_2$ and 400 µM K_4PP_i. The enzymatic reaction was allowed to proceed for 5 minutes at 70 °C. In subsequent assays, NaCl and K_4PP_i were

interchanged in reaction buffer and trigger solution to test Na^+ activation with concentrations adjusted to maintain identical reaction conditions after mixing. Negative controls had no NaCl added to the reaction buffer. The effect of additional components present in the crystallisation condition was tested by supplementing the reaction mixture with 175 mM KCl and/or 26% *v/v* PEG400. Finally, the activity of protein in the optimised vapour diffusion crystallisation condition (60 mM Tris-HCl pH8, 100 mM KCl, 3 mM $MgCl_2$, 175 mM KCl, 26% *v/v* PEG400, 400 µM K_4PP_i) was tested at 20 °C upon activation by supplementing the reaction condition with 20 mM NaCl. The reaction time was extended to 20-1200 minutes to compensate for the reduced hydrolysis rates at below-optimum temperatures and reach the assay sensitivity limit for 12-molybdophosphoric acid detection. 60-240 minutes of incubation gave a linear P_i release with time and the upper limit was chosen for all subsequent assays at 20°°C. Due to the absence of lipids in crystallisation conditions, the activity of non-relipidated sample was tested by diluting OGNG-solubilised protein 200x in reaction buffer that was supplemented with 1.25x CMC OGNG prior to mixing with NaCl trigger solution. After assay development and absorbance measurements at 860 nm the amount of released P_i, specific protein activity and catalytic turnover were calculated. The SEM was obtained from three technical repeats.

2.8.2. Crystal soaking experiments

Time-resolved crystallography experiments were conducted by manual soaking of vapour diffusion crystals that were grown in the absence of Na^+ but in the presence of PP_i (2.6.6) in a Na^+-containing trigger solution (60 mM Tris-HCl pH 8.0, 26% *v/v* PEG400, 175 mM KCl, 2.4 mM $MgCl_2$, 2 mM K_4PP_i, 20 mM NaCl) to initiate the enzymatic *in crystallo* reaction. The reaction was stopped by flash cooling in liquid nitrogen after different soaking times (t=0, 1, 2, 5, 10, 60, 300, 600, 3600 seconds). Crystallisation wells were re-sealed if the soaking time exceeded 60 seconds to minimise evaporation. For each timepoint up to 5 crystals were prepared and diffracted at microfocus beamlines P14-I and P14-II of the Deutsches Elektronen-Syncrhotron (DESY).

2.8.3. Structure solution and refinement

Tm-PPase crystals grown in batch experiments were transferred on a hit-and-return (HARE) silicon support chip using its designated loading platform (192) at beamline P14-II (DESY). The diffraction experiment was carried out at room temperature and the collected data were integrated using XGANDALF (193) of the CrystFEL suite (194) by Dr. David von Stetten (EMBL Hamburg) to obtain the crystal geometry.

Tm-PPase crystals that were grown by vapour diffusion and processed for time-resolved structural studies by soaking (2.8.2) were diffracted at beamline P14-I at DESY. The

X-ray diffraction experiments were carried out at 100 K and the collected data was processed in XDS (180) or Xia2/DIALS (195). This was followed by anisotropy correction using the StarAniso webserver (182) and molecular replacement in Phaser (181) with the *Tm*-PPase:CaMg (PDB: 4AV3) structure as a search model. The similarity between unit cells of the collected datasets was analysed in BLEND (196) and clusters of combined datasets of the same or different time-points (t=0-300 s) were synthesised if the linear cell variation was below 3% and the space group and active site status (occupied *versus* not-occupied) identical.

The single best non-activated structure (reference) and the combined t=0-300 s structure (Cluster A) were subject to several rounds of refinement using phenix.refine (183) and manual modelling in Coot (184). After an initial round of rigid-body refinement with grouped B-factors, torsion angle NCS restraints were applied to further reduce the number of parameters in refinement alongside optimised X-ray/B-factor and X-ray/stereochemistry weighting by phenix. In the final refinement rounds, TLS was applied as well. The t=0-300 s *Tm*-PPase structure was then used as a search model for molecular replacement of combined datasets that were collected after the same time of Na^+-activation. Subsequent refinements followed a similar protocol but were limited to a single round of 5 cycles, which was sufficient to check for changes of the overall helix geometry at the active site or ion gate. Additional secondary structure restraints were applied in the refinement of the low-resolution t=600 s and t=3600 s *Tm*-PPase structure to maintain realistic geometry.

2.8.4. Structural comparison and analysis

Structural alignments and rmsd/C_α calculation were done in PyMol 2.2.3 (186). The selection of atoms to be aligned is described in the text.

Chapter 3. Towards more rational approaches of membrane protein stabilisation

3.1. Introduction

As detailed in 1.1.2, the use of detergents has arguably been the most common and successful strategy for integral membrane protein extraction and solubilisation in the past, but comes at the cost of stripping off native lipids important for protein stability, organisation and function (12,197,198). Even in more “membrane-like” environments such as lipid nanodiscs or SMALPs some membrane proteins are intrinsically unstable once extracted out of the lipid bilayer (28,35). By engineering robust protein variants, screening for different detergents or the addition of stabilising agents, intrinsic instability can be compensated for.

3.1.1. IMPROvER

Protein engineering, particularly the introduction of point-mutations, is a common approach for stabilisation. However, it often renders the protein inactive by rigidifying flexible protein regions and restricting the accessible conformational space (41). Although potentially beneficial for the structural characterisation, conformational stabilisation/trapping often hinders functional studies and time-resolved experiments that could yield valuable insights into complex dynamic processes. Moreover, finding point mutations that improve protein expression, purification or “crystallisability” by systematic or random screening (40) is not trivial and a very labour-intensive and lengthy process (1.1.2).

IMPROvER was developed by former group member Dr. Steven Harborne as a tool for the rational and fast selection of stabilising mutations in α-helical membrane proteins to address these issues (173). It only requires sequence information as input. A set of comparative homology models based on published structural data can be supplied alongside the protein sequence to boost its performance. Additionally, a list of critical residues excluded for mutagenesis can be supplied to bias the selection pool towards active conformations that are suitable for functional or time-resolved biophysical characterisation (173).

IMPROvER combines three modules employing different strategies to identify stabilising mutations: deep-sequence, model-based and data-driven (173). The first module (deep-sequence) makes no prior assumptions about the structure. Large sequence alignments (>8000 sequences) across evolutionarily distant homologues yield information about the natural frequency of a given amino acid at a specific position and

suggest ranked changes in an “odd-ones-out” type analysis (173). Covariance is also taken into account, which implies structural relationships without the requirement for models (173). Sites selected by the deep-sequence approach can improve protein stability directly or affect other aspects of protein biochemistry such as protein expression, folding kinetics or catalytic activity. The second module (model-based) screens for free energy changes (ΔG) upon unfolding during *in silico* thermal melting scans that are performed on homology models (173). Stabilising amino acid substitutions are identified in a comparative analysis of wild-type to variant protein (ΔΔG) (173). They do not rely on accurate prediction of ΔG but only depend on the ability of energy scoring functions to produce correct trends (*i.e.* whether a mutation is stabilising, neutral or destabilising) (199). To limit the influence of poorly modelled protein regions, a stabilising change predicted in multiple models is ranked higher than a stabilising change predicted in only one of the provided models (173). The third module (data-driven) applies general rules extracted from several GPCR stabilisation campaigns with ~2000 individual data points (200–202) on novel protein targets by the means of a scoring matrix (173). The scoring matrix considers amino acid identity, topology, conservation, lipid contact prediction and helix contact prediction and weights factors highest that were most associated with stabilising changes and least associated with destabilising changes (173). These “rules” are then applied to a novel target to predict stabilising mutations (173).

The final output of IMPROvER is a list of ranked mutations with a normalised IMPROvER score from each module and associated primer sequences for site-directed mutagenesis (173). The simplicity of this bioinformatic approach allows processing of an input query on a single 2.4 GHz central processing unit in less than 72 hours (dependent on the length of protein sequence), removing the requirement for access to high-performance computing facilities (173).

3.1.2. A high-throughput lipid screen for stabilising membrane proteins

The reintroduction of lipids to detergent-solubilised proteins is an alternative stabilisation method that proved crucial for the characterisation of a broad range of targets such as GPCRs (51), transporters (50), ion channels (48), ion pumps (47) and electron transport complexes (49). The experimental burden associated with finding stabilising lipids is lower compared to identifying stabilising point mutations; however, it is still time-consuming and laborious as it also relies on a trial-and-error process. Moreover, screening a large lipid space is very expensive as each lipid needs to be purchased individually and prevailing commercial aliquot sizes are much larger than what is required for experimental evaluation. These issues were tackled in a collaborative effort with

Cristina Cecchetti, Claudia Stohrer and our industrial RAMP-ITN partner Molecular Dimensions by developing a microplate lipid screen.

3.1.3. Aims and strategy

As part of RAMP, my contribution to the rationalisation of membrane protein crystallisation focused on developing and testing novel tools for protein stabilisation by protein engineering (IMPROvER) or relipidation (RAMP lipid screen).

I evaluated IMPROvER's performance and experimentally established of a scalable screening strategy based on *Cl*-PPase as test protein, a K^+-dependent Na^+,H^+-PPase for which no structural data is available yet. Previous efforts towards a *Cl*-PPase structure were hampered by its tendency to aggregate at the concentrations required for crystallisation (4.1.1). For a more robust assessment of IMPROvER's prediction power and demonstration of its broad applicability, the equilibrative nucleoside transporter 1 (hENT1) and human parathyroid hormone 1 receptor ($hPTH_1R$) were stabilised in complementary studies by group members Jessica Boakes and James Henderson, respectively, based on initial findings of the *Cl*-PPase stabilisation campaign (173). In future, IMPROvER can be used to significantly lower the experimental burden of protein stabilisation campaigns and accelerate the research of challenging high-impact protein targets as general-purpose tool for the more rational selection of point-mutations.

The identification of beneficial lipids was streamlined for cost-effectiveness and high-throughput applicability by developing a multi-purpose microplate screen, the RAMP lipid screen. This screen contains sufficient amounts of 31 different lipids or lipid mixtures to sample a broad range of lipids at affordable cost and with minimal effort. I assessed its performance and practicability on three membrane proteins under investigation by the RAMP network: a bacterial pyrophosphatase (*Tm*-PPase), a fungal purine transporter (UapA) and a human GPCR ($A_{2A}R$). These targets represent a set of challenging membrane proteins that can be purified in suitable amounts and have different folds (from 7 to 16 transmembrane helices), topologies and modes of action. In future, the RAMP lipid screen will hopefully facilitate the structural and functional analysis of stable and physiologically relevant protein samples as robust and simple platform for the rapid identification of beneficial lipids.

3.2. Results

3.2.1. *Cl*-PPase as test case for IMPROvER

3.2.1.1. Computational prediction of stabilising point mutations

Stabilising point mutations of *Cl*-PPase were predicted by employing all three modules, (deep-sequence, model-based, data-driven) of the IMPROvER software package. 96 residues, all from the top 15% of a ranked list compiled by IMPROvER, were taken forward to experimental validation (Figure 3.1). Other test proteins for IMPROvER had more ($hPTH_1R$: 10%) or less (hENT1: 40%) strict selection cut-offs for ranked lists applied to assess if there is a correlation between rank and stabilisation (173). *Cl*-PPase was cloned into the pDDGFP2 yeast expression plasmid with a N-terminal His-tag and sfGFP attached. This construct design was previously used in similar studies for the rapid identification of ideal expression and purification conditions or stabilising point mutations in membrane proteins (176,203). Moreover, *Tm*-PPase and *Pa*-PPase are known to express well in *S. cerevisiae* (204).

The prediction set was filtered against a list of residues critical for the structural and functional integrity of M-PPases (Table S5 , Appendix). This excluded 24 highly ranked sites from mutagenesis to bias the prediction towards active protein variants (Figure 3.1A). There was only little overlap between different prediction approaches with 13 sites suggested for mutagenesis by more than one module (Figure 3.1B). In case of overlapping predictions, only the higher ranked amino acid was selected for mutation to cover a larger sequence space, even if the suggested change differed (*e.g.* deep-sequence: $I^{6.60}V$, structure-based: $I^{6.60}W$). After several iterations of site-directed mutagenesis using inverse PCR with back-to-back mutagenesis primers, 45 of 96 variants were successfully introduced. The variants provided a random sampling of the selected sites with 15 mutations from the deep-sequence, 15 mutations from the model-based and 15 mutations from the data-driven module (Figure 3.1A, Table 3.1).

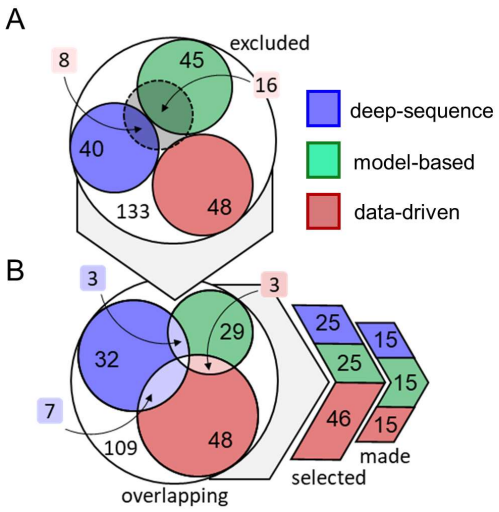


Figure 3.1: Prediction of stabilising mutations by IMPROvER. (A) Number of putatively stabilising mutations predicted by each module. The overlap with critical residues is shown by a black circle. (B) Number of putatively stabilising mutation after exclusion of critical residues with the prediction overlap between modules shown. 96 mutations were selected, all ranked within the top 15%. Figure adapted from Harborne et al. (172).

3.2.1.2. Mutagenesis and experimental evaluation of the predictions

Wild-type and variant *Cl*-PPase expression in *S. cerevisiae* was confirmed by whole-cell GFP detection (Figure S1, Appendix), but not analysed in detail because I was primarily screening for stabilising mutations and did not consider secondary effects such as improved protein expression or folding kinetics. The *Cl*-PPase expression level of wild-type protein was ~0.1 mg L^{-1} culture based on its GFP fluorescence signal, too little for structural studies but sufficient for GFP-based thermostability screening. GFP fluorescence allowed variant evaluation from detergent-solubilised crude membrane extract without the need to purify, and followed a protocol by developed by Ashok and co-workers (205).

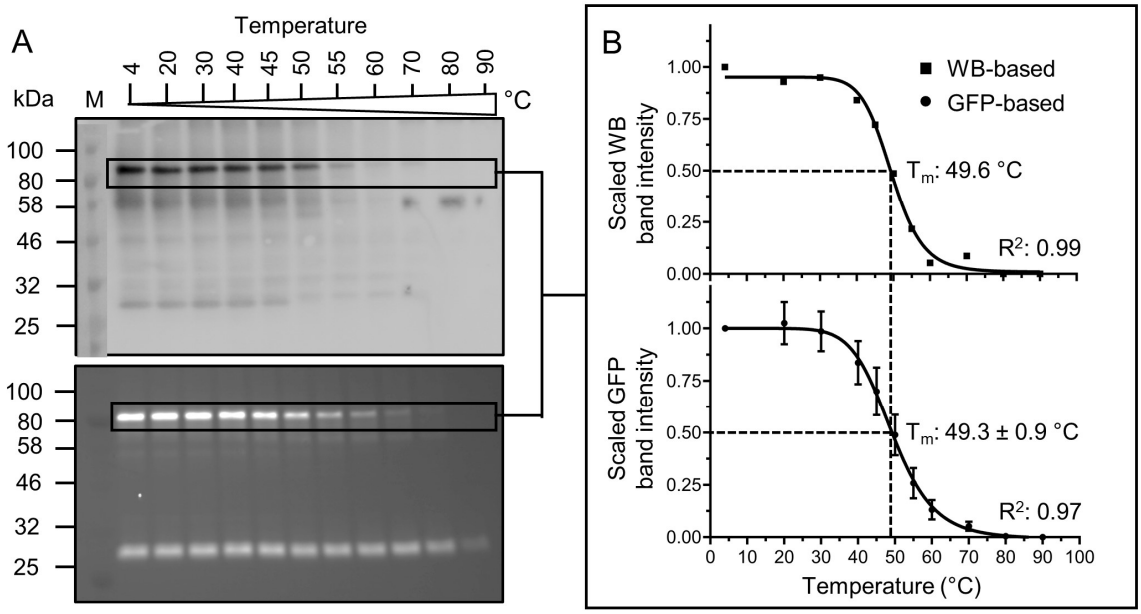


Figure 3.2: Western-Blot and GFP-based ten-temperature thermostability assay of wild-type *Cl*-PPase for T_m determination. (A) T_m determination based on anti-His Western-Blot (upper panel) or in-gel GFP fluorescence (lower panel). (B) Black box highlights full-length *Cl*-PPase bands used in densiometric analysis to obtain melting curves by plotting the fluorescence intensity against the temperature challenge. Dotted lines highlight the temperature at which half of the protein population is unfolded (T_m). The GFP signal of melting curves is normalised to the intensity of the reference sample incubated on ice. Error bars denote SEM.

Wild-type and variant *Cl*-PPase samples were heat challenged and the GFP signal monitored to assess the resistance of the protein to precipitation at elevated temperatures as a measure of thermostability (Figure 3.2). After heating, any precipitation was removed by centrifugation and the remaining protein was quantified by densiometric analysis of the in-gel fluorescence associated with full-length *Cl*-PPase. The apparent melting temperature (T_m) at which half of the protein population is unfolded was determined in ten-temperature thermostability assays (4-90 °C) to quantify the degree of stabilisation. Originally, Ashok and co-workers visualised the protein of interest by Western-Blotting instead of GFP fluorescence detection, but the latter worked equally well and further reduced the workload when tested on wild-type *Cl*-PPase (Figure 3.2A).

The wild-type T_m, determined to be 49.3 ±0.9 °C, was used as reference in one-temperature thermostability assays and analysis (Figure 3.2B).

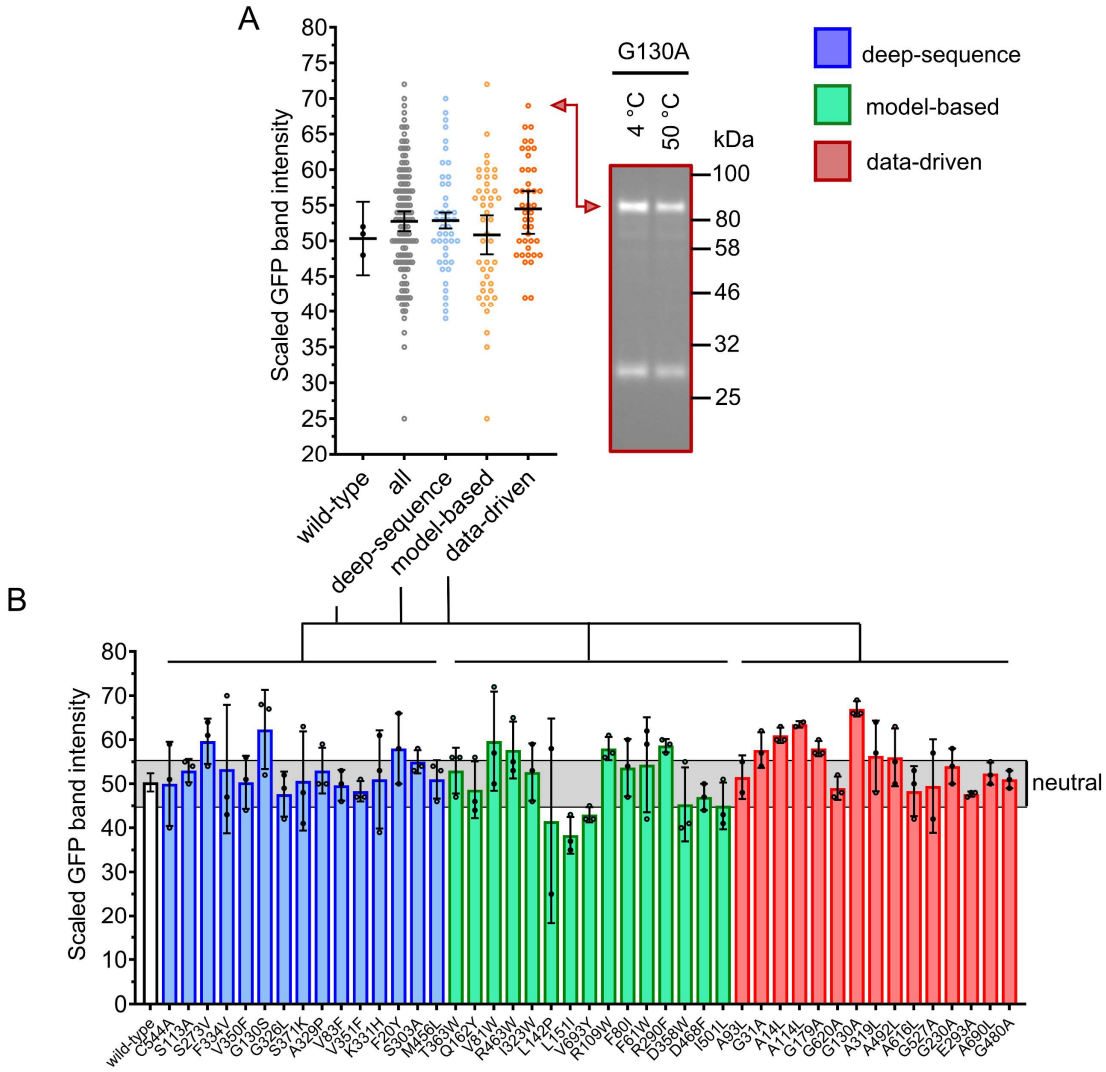


Figure 3.3: One-temperature thermostability assay of variant *Cl*-PPase. (A) Evaluation of prediction modules in one-temperature thermostability assays. The mean GFP fluorescence survival is indicated by a horizontal line and error bars denote the 95% CI of all data-points in the respective subgroup. An exemplary GFP-visualised SDS-PAGE gel that was used for analysis of the one-temperature thermostability assay is shown for one data point. (B) Evaluation of individual mutations in one-temperature thermostability. All data points are plotted as circles and error bars denote the SEM. The selection cut-off for recording full melting curves is visualised by grey background shading. Putative stabilisers extend beyond the background shading. Figure adapted from Harborne *et al.* (173).

The one-temperature assays were used in a semi-quantitative screening for an initial fast evaluation of *Cl*-PPase variants (Figure 3.3). Here, the fluorescence survival after a single temperature challenge at the T_m of wild-type *Cl*-PPase was measured to make a pre-selection of variants for recording full melting curves (Figure S2, Appendix). The cut-off for recording full melting curves was set at ≥55% fluorescence survival, which corresponds to the 95% confidence interval of all data points (Figure 3.3A). Consequently, only variants likely to stabilise the most were selected, *i.e.* 95% confidence that the selected variant confers above average stabilisation.

Based on this stringent selection criteria, 14 of 45 mutations were selected for further characterisation (Figure 3.3B, Table 3.1). To confirm initial screening and quantify the degree of stabilisation, full melting curves were recorded for these variants in ten-temperature thermostability assays (Figure 3.4A-D). Additionally, the melting curves of 3 destabilising and 4 neutral mutations were also recorded to evaluate the screening strategy. Variants were then classified as stabilising or destabilising only if the ΔT_m was ≥ ±1.3 °C compared to wild-type *Cl*-PPase, a cut-off determined by the average SEM of all ten-temperature thermostability assays.

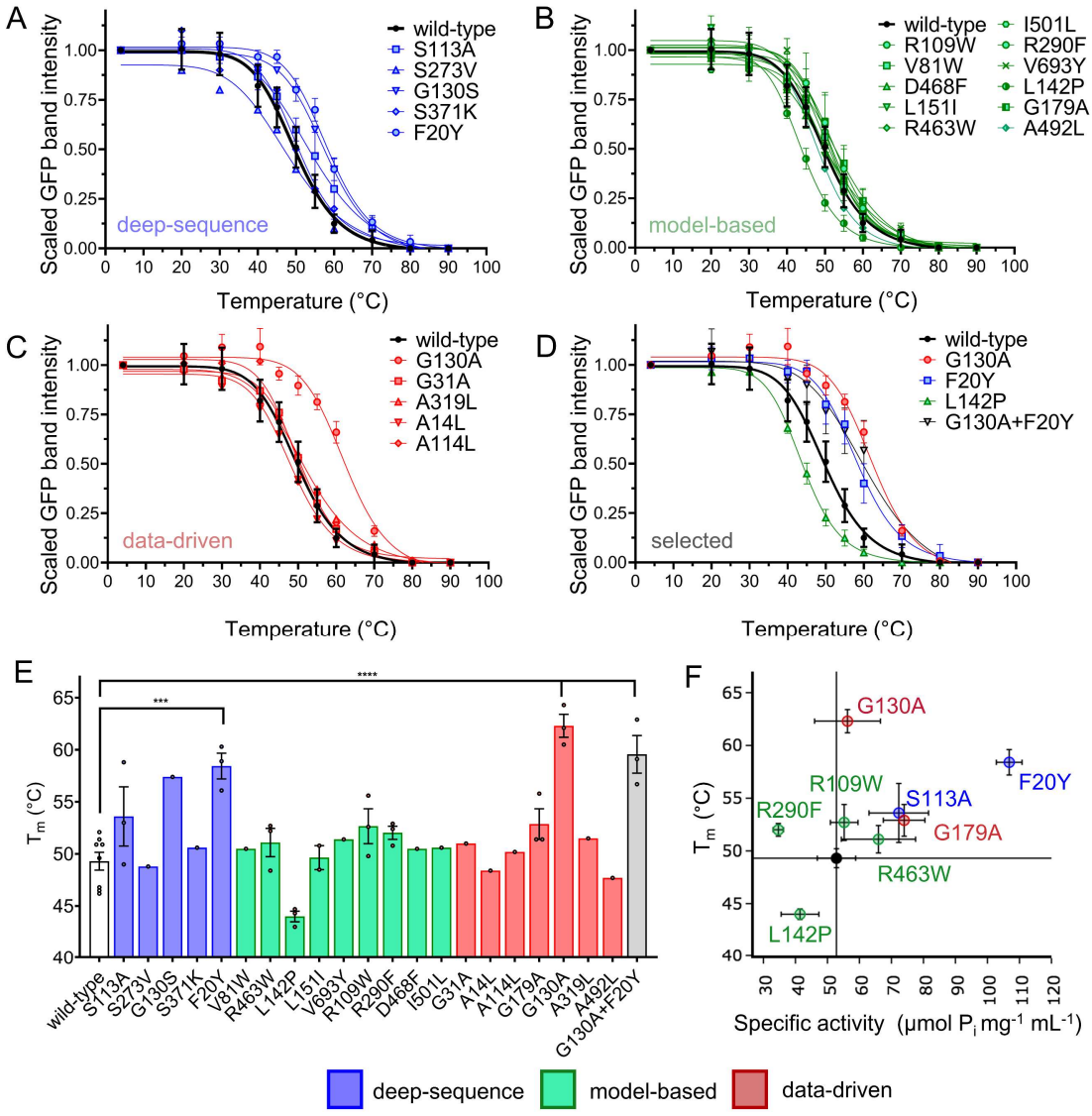


Figure 3.4: Ten-temperature thermostability assay of variant *Cl*-PPase. (A-C) Full melting curves recorded for variants predicted by the (A) deep-sequence, (B) model-based and (C) data-driven module of IMPROvER. (D) Selected subset of melting curves. Colouring according to prediction module with deep-sequence in blue, model-based in green and data-driven in red. Wild-type melting curves are shown in black. The GFP signal of melting curves is normalised to the intensity of the reference sample incubated on ice. (E) T_m determination from full-melting curves. Individual values from repeated measurements are plotted as circles. Statistically significant differences are indicated by *** ($p<0.0004$) or **** ($p<0.0001$). (F) T_m of the top stabilising variants and a destabilising variant (L142P) plotted against the specific activity. Stabilising variants that retain or improve PP_i hydrolysis activity cluster in the top right quadrant. Figure adapted from Harborne *et al*. (173).

Table 3.1: IMPROvER mutations characterised by one- and ten-temperature thermostability assays.Adapted from Harborne *et al.* (173).

Module	Score[a]	Mutation	GFP fluorescence survival[b] (%)	Error[c] (±%)	n[d]	T_m[e] (°C)	Error[c] (±°C)	n[d]	Identity[f] (%)	Effect
n/a	n/a	Wild-type	50.3	1.2	3	49.3	0.9	8	n/a	n/a
deep	1.00	C544A	50.0	5.5	3	-	-	-	70 (A)	neutral
deep	1.00	S113A	53.0	1.5	3	53.6	2.8	3	35 (A)	stabilising
deep	1.00	S273V	59.7	3.0	3	48.8	-	1	42 (L)	neutral
deep	1.00	F334V	53.3	8.4	3	-	-	-	19 (I)	neutral
deep	0.99	V350F	50.3	3.5	3	-	-	-	11 (X)	neutral
deep	0.97	G130S	62.3	5.2	3	57.4	-	1	100 (G)	stabilising
deep	0.97	G326L	47.7	3.0	3	-	-	-	14 (L)	neutral
deep	0.97	S371K	50.7	6.5	3	50.6	-	1	16 (E)	neutral
deep	0.97	A329P	53.0	3.0	3	-	-	-	10 (D)	neutral
deep	0.96	V83F	49.7	2.0	3	-	-	-	19 (X)	neutral
deep	0.96	V351F	48.3	1.3	3	-	-	-	15 (I)	neutral
deep	0.95	K331H	51.0	6.4	3	-	-	-	10 (X)	neutral
deep	0.95	F20Y	58.0	4.6	3	58.4	1.2	3	24 (Y)	stabilising
deep	0.95	S303A	55.0	1.5	3	-	-	-	36 (A)	neutral
deep	0.94	M456L	51.0	2.5	3	-	-	-	37 (L)	neutral
model	1.00	T363W	53.0	3.0	3	-	-	-	46 (V)	neutral
model	0.99	Q162Y	48.7	3.7	3	-	-	-	4 (X)	neutral
model	0.99	V81W	59.7	6.5	3	50.5	-	1	27 (T)	stabilising
model	0.99	R463W	57.7	6.0	3	51.1	1.3	3	40 (R)	stabilising
model	0.99	I323W	52.7	3.8	3	-	-	-	15 (T)	neutral
model	0.98	L142P	41.5	16.5	2	44.0	0.5	3	41 (L)	destabilising
model	0.98	L151I	38.3	2.4	3	49.7	1.2	2	24 (L)	neutral
model	0.97	V693Y	43.0	1.0	3	51.4	-	1	25 (I)	stabilising
model	0.97	R109W	58.0	1.5	3	52.7	1.7	3	95 (R)	stabilising
model	0.96	F80I	53.7	3.8	3	-	-	-	10 (X)	neutral
model	0.95	F61W	54.3	6.2	3	-	-	-	18 (F)	neutral
model	0.95	R290F	58.7	0.9	3	52.0	0.6	3	22 (Y)	stabilising
model	0.95	D358W	45.3	4.8	3	-	-	-	14 (Y)	neutral
model	0.94	D468F	47.0	1.7	3	50.5	-	1	100 (D)	neutral
model	0.94	I501L	45.0	3.1	3	50.6	-	1	28 (V)	neutral
data	0.99	A93L	51.5	3.5	2	-	-	-	74 (G)	neutral
data	0.98	G31A	57.7	2.3	3	51.0	-	1	52 (G)	stabilising
data	0.96	A14L	61.0	1.0	3	48.4	-	1	36 (G)	neutral
data	0.95	A114L	63.5	0.5	2	50.2	-	1	70 (A)	neutral
data	0.91	G179A	58.0	1.0	3	52.9	1.5	3	74 (G)	stabilising
data	0.90	G620A	49.0	1.5	3	-	-	-	100 (G)	neutral
data	0.88	G130A	67.0	1.0	3	62.3	1.1	3	100 (G)	stabilising
data	0.85	A319L	56.3	4.6	3	51.5	-	1	14 (X)	stabilising
data	0.85	A492L	56.0	3.8	3	47.7	-	1	62 (A)	destabilising
data	0.84	A616L	48.3	3.3	3	-	-	-	20 (A)	neutral
data	0.79	G527A	49.5	9.0	3	-	-	-	42 (A)	neutral
data	0.75	G230A	54.0	2.3	3	-	-	-	100 (G)	neutral
data	0.71	E293A	47.7	0.3	3	-	-	-	19 (E)	neutral
data	0.70	A690L	52.3	1.5	3	-	-	-	55 (A)	neutral
data	0.68	G480A	51.0	1.2	3	-	-	-	100 (G)	neutral

[a]IMPROvER score used for ranking sites
[b]Remaining in-gel GFP fluorescence after heat challenge in one-temperature thermostability assay
[c]For n>1: SEM shown. For n=1, error estimate is ±1.3 °C based on average SEM in ten-point assay.
[d] number of experimental repeats
[e]Calculated by best fit of a four-parameter does-response curve (varaible slope) by non-linear least-squares fitting in GraphPad prism 9.0
[f]Sequence identity and consensus set in parentheses based on a blast(p) search for similar sequences to *Tm*-PPase in Geneious R11

12 variants of the pre-selection were confirmed to be stabilising, 7 variants were neutral and only 2 variants were destabilising (Figure 3.4E, Table 3.1). This translates to an overall success rate for identifying stabilising mutations by IMPROvER of 27%, which is 3-4x better than randomly selecting mutations in the GPCR training dataset (~7.6%) that was used to derive the data-driven scoring matrix (173). The best variant, G130A, increased the T_m compared to wild-type *Cl*-PPase by +13.0 ±1.4 °C followed by F20Y

with +9.1 ±1.5 °C (Figure 3.4D-E). Combining the two best variants G130A and F20Y did not improve thermostability any further (ΔT_m= 10.0 ±2.0 °C) (Figure 3.4D-E). Interestingly, all top stabilisers showed comparable or elevated hydrolytic activity levels to wild-type *Cl*-PPase in membranes (Figure 3.4F), which highlights the benefit of filtering the prediction set against a list of critical residues to select a pool of active variants. The individual modules were similarly powerful with success rates of 20%, 33% and 27% for the deep-sequence, model-based and data-driven approach. Rapid screening by one-temperature assays also proved successful at pre-selecting putative stabilisers with false positive and false negative rates of ~29%, respectively, but ~5x increased throughput.

3.2.2. Lipid screen development and testing

The development of the RAMP lipid screen was split into two parts. The first part was the rational selection of lipids, design and production of the screen. The second part was the development of a high-throughput screening protocol and evaluation of the new screen on a range of membrane proteins.

3.2.2.1. Rationale and screen design

The selection of lipids was based on an extensive literature search utilising the LipidMAPS (https://www.lipidmaps.org/) and the PDB (https://www.rcsb.org/) to extract information about the identity and frequency of lipids specifically bound to membrane proteins from native mass spectrometry data, lipidomics analysis and high-resolution structures. Additionally, the materials and methods section of selected structural biology publications were studied to pinpoint lipids commonly used in protein purification, cryo-EM grid preparation or crystallisation. Cristina Cecchetti and I chose a total of 23 different lipids of natural and synthetic origin and 8 defined lipid mixtures for the screen (Table 3.2). These lipids are known to affect protein stability, function, organisation or structure in various membrane proteins (Table 3.2). We specifically selected a broad range of lipid types to sample a large lipid space including phosphatidylethanolamines (PEs), phosphatidylcholines (PCs), phosphoglycerides (PGs), phosphatidylserines (PSs), phosphatidic acids (PAs), cardiolipins (CLs), sphingolipids (SLs), monoacylglycerols (MAGs) and sterols.

The two major components of prokaryotic and eukaryotic membranes, PEs and PCs, are broadly used in functional and structural studies of all types of membrane proteins (32,206–211). Protein-lipid interactions with less abundant lipids such as PGs, PSs and PAs, all negatively charged, have been demonstrated to be of importance for the oligomerisation and function of a several transporters and channels (212–214). Sterols, particularly cholesterol (CHL), and the more soluble salt cholesterol hemisuccinate (CHS) are known to bind to a range of mammalian membrane proteins

including GPCRs (34) and transporters (215–217). CLs were also shown to associate with membrane proteins, often respiratory complexes (218) but also transporters (31). The screen also includes MAGs, which play an important role in structural studies of membrane proteins as host lipids for LCP (1.1.3). The natural and defined lipid mixtures supplied with the screen have previously aided the study of stable and physiologically more relevant protein samples, for example ion channels (219–222). Some lipids were not included in the screen despite their biological importance to ensure a reasonable pricing and practical shelf life of the screen (*e.g.* phosphatidyl inositols (PIs)).

A first test batch of screens was prepared in a 96-well format in glass-coated microplates, which are resistant to the organic solvents commonly used to dissolve lipids. Each lipid or lipid mixture was dispensed in triplicate (Table 2.3). The amount of dispensed lipid, 0.3 mg per well, was chosen to be suitable for a broad range of applications. These include functional analysis (*e.g.* in liposomes) (25), nanodisc reconstitution (28), stability screening (45,46) and HiLiDe crystallisation (68), of which the latter typically requires the most lipid amount. After the production of a first batch of screens, I assessed its potential as high-throughput platform for the identification of stabilising lipids on three membrane proteins: $A_{2A}R$, UapA and *Tm*-PPase.

3.2.2.2. High-throughput protein relipidation and stability screening

All lipids supplied in the screen were solubilised in detergent ($A_{2A}R$, UapA: DDM, *Tm*-PPase: OGNG) prior to mixing with sample for protein relipidation. However, most lipids did not solubilise completely, which resulted in an unknow final lipid concentration. *Tm*-PPase was purified following the “hot-solve” protocol, yielding good-quality sample after IMAC (Figure 3.5). Purified $A_{2A}R$ and UapA were supplied by Claudia Stohrer and Cristina Cecchetti, respectively (Figure 3.5). The laddering seen for $A_{2A}R$ is commonly observed and likely a results of glycosylation (223), which should not interfere with stability measurements. $A_{2A}R$ served as positive control in combination with CHS and CHL, for which its interactions have been characterised extensively (216). Additionally, high-affinity ligands ($A_{2A}R$: ZM241385 (ZM), UapA: Xanthine, *Tm*-PPase: IDP) were used as secondary positive control to verify

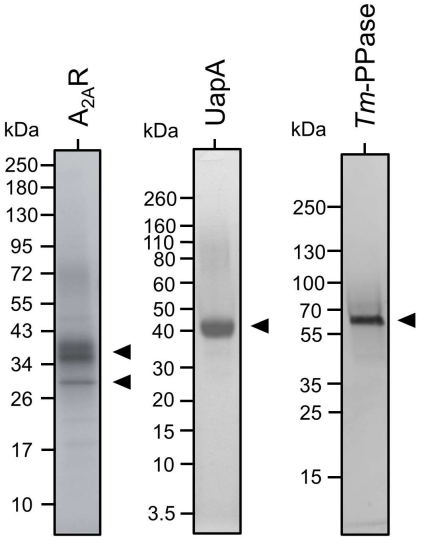


Figure 3.5: Protein purification for DSF stability assays SDS-PAGE analysis of the elution fractions of IMAC/SEC purification of three targets. Protein bands were visualised by Coomassie staining and full-length protein bands are highlighted by a black arrow. Figure adapted from Cecchetti et al. (2).

the detection of stabilising effects by differential scanning fluorimetry (DSF).

DSF detects changes of the intrinsic protein fluorescence caused by the reorientation of the local tryptophan (and tyrosine) residue environment upon thermal or chemical unfolding (224). I used a Prometheus NT.48 nanoDSF system to record melting curves (thermal unfolding) and extract the T_m of test proteins supplemented with (relipidated sample) or without (reference sample) lipids. The protein concentration used in nanoDSF measurements was chosen to give sufficient fluorescence signal with minimal sample consumption (~0.75-1.5 mg protein for screening in triplicate). Changes of the intrinsic tryptophan fluorescence intensity were monitored at 330 nm (F_{330}) and 350 nm (F_{350}) throughout sample heating and showed clear transitions for $A_{2A}R$ at 43.3 ±0.2 °C and *Tm*-PPase at 81.8 ±0.3°C, respectively (Figure 3.6A/C, Table 3.2). For UapA, the $F_{350:330}$ ratio was more informative and showed a clear transition at 46.3 ±1.1 °C (Figure 3.6B, Table 3.2). The $F_{350:330}$ is more sensitive to small changes in the tryptophan environment as it is less susceptible to fluorescence backgrounds and highlights emission shifts (*e.g.* “red-shift” or “blue-shift”). Following the previous classification scheme for evaluation of IMPROvER (3.2.1.2), stabilising effects were defined by a T_m increase upon lipid addition exceeding the individual and average SEM ($A_{2A}R$: 0.4°°C, *Tm*-PPase: 0.3 °C UapA: 0.8°°C) of all measurements for T_m determination (Figure 3.7).

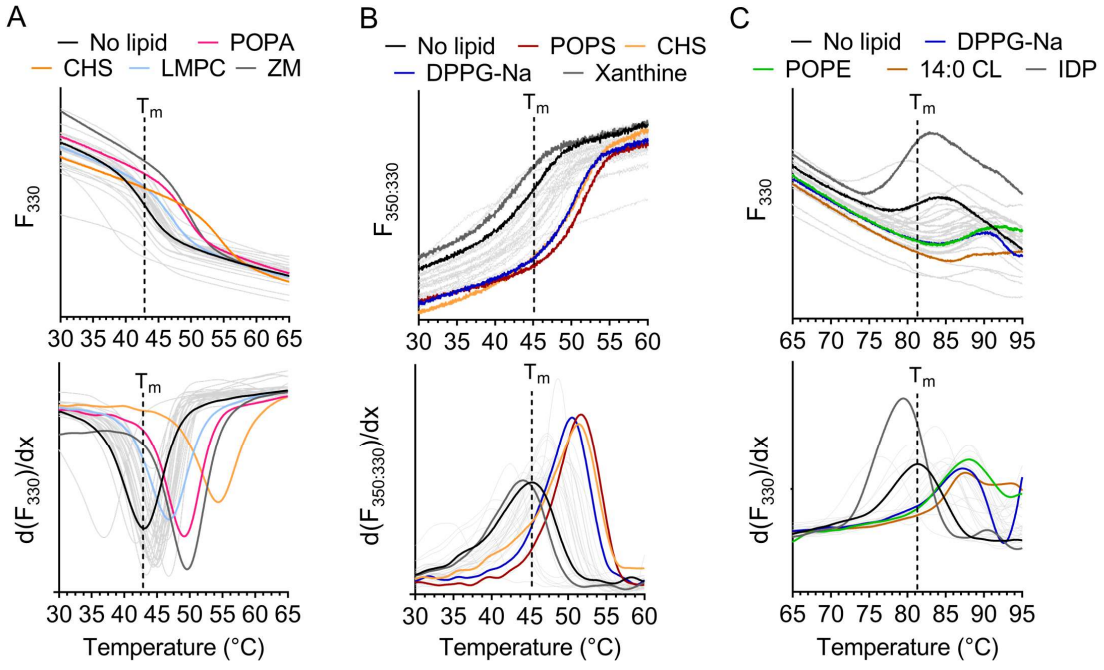


Figure 3.6: nanoDSF thermostability assays for lipid screen evaluation. (A) Raw fluorescence traces (upper panel) and first derivate thereof (lower panel) for melting scans done for $A_{2A}R$, UapA and *Tm*-PPase plotted in grey. Data from no lipid, ligand and top stabilising lipids is highlighted by thick coloured lines. The inflection point of the unfolding transition (T_m) for the no lipid reference sample is indicated by a dashed line.

Intriguingly, MAGs were consistently destabilising (or neutral) for all three proteins despite their success in the structure determination of membrane proteins by LCP.

For $A_{2A}R$, a total of 22 lipids were stabilising, 7 were neutral and 2 were destabilising of the 31 different lipids or lipid mixtures tested (Figure 3.7, Table 3.2). The average degree of stabilisation was +2.4 ±0.3 °C with the best lipid, CHS, stabilising by +10.8 ±0.2°C, followed by 16:18 PA with +5.8 ±0.2°C and LMPC with +3.6 ±0.2°C (Table 3.2). UapA was also stabilised by +2.4 ±1.1 °C on average, but fewer stabilising lipids were identified (11 stabilising, 11 neutral, 9 destabilising) (Figure 3.7, Table 3.2). POPS, CHS and DPPG had the most positive effect on protein stability with a ΔT_m of +5.1 ±1.1 °C, +4.2 ±1.1 °C and +3.9 ±1.1 °C, respectively (Table 3.2). Beside MAGs, PAs and CLs were the most destabilising, lowering the T_m by up to -11.7 ±1.1 °C (10:PA). *Tm*-PPase was stabilised by almost every lipid with an average degree of stabilisation of +5.2 ±0.4 °C (Figure 3.7, Table 3.2). The biggest increase in T_m was observed for POPE, DPPG and 14:0 CL with +7.4 ±0.6 °C, +7.4 ±0.4 °C and +7.3 ±0.4 °C, respectively (Table 3.2). Finally, ligand binding was confirmed stabilising in case of $A_{2A}R$ ($\Delta T_{m,ZM}$: +5.3 ±0.5 °C), but not for UapA and *Tm*-PPase. Xanthine (UapA) and IDP (*Tm*-PPase) affected protein stability negatively ($\Delta T_{m,xanthine}$: -2.4 ±1.2 °C, $\Delta T_{m,IDP}$: -1.1 ±0.4 °C) (Figure 3.7, Table 3.2). All stability data was recorded in triplicate using only one screen within 1 day, which highlights the cost-effectiveness of the RAMP lipid screen and the simplicity of the screening strategy.

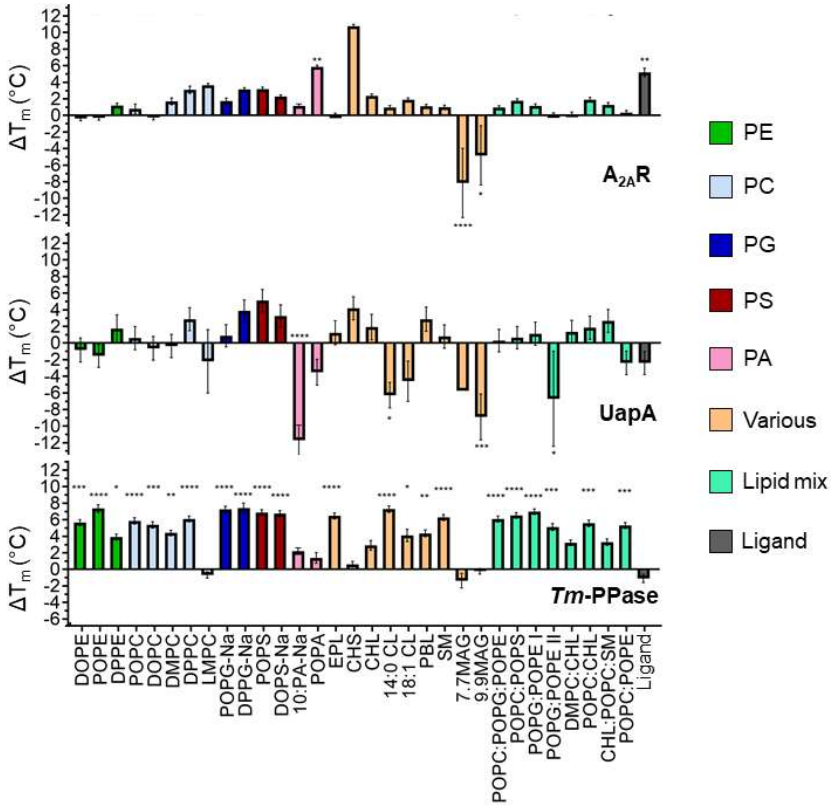


Figure 3.7: Summary of nanoDSF stability testing for lipid screen evaluation.ΔT_m of detergent-solubilised to relipidated sample shown for $A_{2A}R$, UapA and *Tm*-PPase. The ligand for $A_{2A}R$ (ZM241385), UapA (Xanthine) and *Tm*-PPase (Na_4IDP) differs. Lipids are sorted and colour coded according to their lipid class. Error bars denote the SEM based on three technical repeats. The asterisks represent the p-value with * (0.0332), ** (0.021), *** (0.002) and **** (0.0001). Figure adapted from Cecchetti *et al.* (172).

Table 3.2: Summary of nanoDSF stability testing for lipid screen evaluation.Adapted from Cecchetti *et al.* (172).

Lipid	$A_{2A}R$			UapA			*Tm*-PPase			Reference
	ΔT_m[a]	error[b]	n[c]	ΔT_m[a]	error[b]	n[c]	ΔT_m[a]	error[b]	n[c]	
no lipid	43.3	0.2	3	46.3	1.1	3	81.1	0.3	3	-
DOPE	43.0	0.1	3	45.5	0.5	3	86.7	0.1	3	(206)
POPE	43.0	0.2	3	44.8	0.4	3	88.4	0.3	3	(32)
DPPE	44.5	0.2	3	48.1	1.0	3	85.0	0.2	3	(206)
POPC	44.2	0.5	3	46.9	0.4	3	86.9	0.2	3	(207)
DOPC	43.1	0.2	3	45.7	0.5	3	86.4	0.2	3	(208)
DMPC	45.0	0.4	3	46.0	0.4	3	85.4	0.1	3	(209)
DPPC	46.4	0.4	3	49.2	0.3	3	87.1	0.1	3	(210)
LMPC	47.0	0.1	3	44.1	3.6	3	80.4	0.2	3	(211)
POPG-Na	45.1	0.3	3	47.2	0.1	3	88.3	0.2	3	(212)
DPPG-Na	46.5	0.1	3	50.2	0.1	3	88.5	0.5	3	(213)
POPS	46.5	0.1	3	51.4	0.1	3	87.9	0.2	3	(214)
DOPS-Na	45.6	0.0	3	49.6	0.1	3	87.8	0.2	3	-
10:PA-Na	44.5	0.1	3	34.6	1.2	3	83.2	0.2	3	(225)
16:0-18:1 PA (POPA)	49.2	0.1	3	42.8	0.8	3	82.4	0.6	3	-
EPL	43.4	0.1	3	47.6	0.5	3	87.5	0.2	3	(219)
CHS	54.1	0.1	3	50.5	0.3	3	81.7	0.1	3	(226)
CHL	45.7	0.1	3	48.2	0.8	3	83.9	0.5	3	(216)
14:0 CL	44.3	0.1	3	40.1	0.7	3	88.3	0.2	3	(31)
18:1 CL	45.2	0.0	3	41.8	2.0	3	85.2	0.7	3	(31)
PBL	44.4	0.1	3	49.2	0.6	3	85.4	0.3	3	(46)
SM	44.3	0.2	3	47.1	0.4	3	87.3	0.2	3	(45)
7.7MAG	35.2	4.2	3	40.6	-	1	79.7	0.8	3	(64)
9.9MAG	38.5	3.6	3	37.4	2.4	3	80.8	0.1	3	(62)
POPC:POPG:POPE (3:1:1)*	44.3	0.1	3	46.6	0.2	3	87.1	0.2	3	(220–222)
POPC:POPS (4:1)*	45.1	0.1	3	47.0	0.1	3	87.6	0.1	3	-
POPG:POPE (3:1)*	44.5	0.1	3	47.5	0.4	3	88.0	0.1	3	-
POPG:POPE (3:1)*	43.4	0.1	3	39.6	5.6	3	86.2	0.3	3	-
DMPC:CHL (2:1)*	43.5	0.1	3	47.7	0.2	3	84.3	0.1	3	-
POPC:CHL (5:1)*	45.2	0.2	3	48.2	0.4	3	86.6	0.2	3	-
CHL:POPC:SM (1.8:1:1)*	44.6	0.2	3	49.0	0.3	3	84.3	0.2	3	-
POPC:POPE (1:1)*	43.7	0.1	3	43.9	0.4	3	86.3	0.2	3	-
ZM241385 (ZM)	48.6	0.5	3	-	-	-	-	-	-	-
Xanthine	-	-	-	44.0	0.5	3	-	-	-	-
IDP	-	-	-	-	-	-	80.0	0.3	3	-

Colour coding relates to lipid effect on protein stability: stabilising (green), neutral (white), destabilising (red)
[a] Calculated by PR.ThermControl software based on F_{330}, F_{350} or $F_{350:330}$ signal
[b] For n>1: SEM shown. For n=1, error estimate is ±0.8 °C based on average SEM of T_m determination by nanoDSF
[d] Number of experimental repeats
[c] Number of experimental repeats
* Molar ratio

3.3. Discussion

Protein stability remains one of the major bottlenecks for investigating membrane proteins and their biophysical characterisation often requires purified sample in relatively large amounts (*e.g.* milligram quantities for X-ray crystallography). Most protocols purify protein in detergent micelles, which are only poor mimics of the lipid bilayer and tend to destabilise membrane proteins, or require an intermediate detergent extraction step, resulting in the loss of most if not all the native lipids important for functional and structural integrity (1.1.2). I contributed to the development and evaluation of two new tools designed to address issues of protein instability and advance membrane protein research, namely IMPROvER and the RAMP lipid screen.

3.3.1. *Cl*-PPase as test case for IMPROvER

IMPROvER proved successful for the *in silico* prediction of stabilising point-mutations in *Cl*-PPase with hit rates about 3-4x better than random selection. This enables screening for stabilising mutations in large protein targets, even in academic settings. Usually, the identification of beneficial point-mutations by scanning-mutagenesis takes 6-12 months for a 300 amino acid protein (227). Here, I stabilised a 700 amino acids protein in just ~3 months following a more rational approach.

IMPROvER is not the first bioinformatic tool predicting stabilising point mutations in membrane proteins, but it is the first tool that is broadly applicable to proteins of different fold, topology and modes of action while outperforming comparable software (203,228-231). Importantly, it only requires sequence information as minimal input, a distinct advantage over many other tools that rely on additional data. For example, Sauer and co-workers developed a combined approach of deep-sequencing and optimal growth temperature analysis to identify amino acid changes characteristic for intrinsically thermostable proteins and transfer them into mesophilic orthologues (228). While this approach is promising for M-PPases, where several thermostable orthologues are known, it cannot predict stabilising mutations in proteins that lack heat resistant family members. Another example are tools like mCSF-membrane that require *a priori* structural knowledge (229). mCSF-membrane employs machine learning based on distance patterns between atoms to represent residue environments, which excludes *de novo* targets for stabilisation (229). However, this and similar approaches will benefit from recent advances in the computational prediction of protein structures with atomic accuracy (see Alphafold 2) (232). Yet other prediction tools are only applicable to certain protein classes (230,231). CompoMug, for example, is capable of *de novo* protein stabilisation, but only works for GPCRs. It utilises sequence information in deep-sequencing and data-driven modules (231) in addition to structural information for machine learning and salt-bridge analysis (231). The combinational prediction approach and some of the individual modules are similar, but not identical, to those of IMPROvER. This reflects in a comparable success rates of 25% and 27% for CompoMug and IMPROvER, respectively (231).

IMPROvERs success rate was shown to increase to up to 40% with a more stringent selection strategy that focused on the top10% of ranked mutations ($hPTH_1R$) as opposed to the top 15% (*Cl*-PPase) or top 40% (hENT1) in complementing studies (173). Overall, ranking by IMPROvER did not directly correlate with the experimentally determined degree of thermostabilisation (Table 3.1). Instead, ranking guided the rational selection of sites that are most likely to be involved in stabilising interactions (the top 10-15%) and dramatically speeds up the workflow.

Most stabilising mutations in *Cl*-PPase increased the T_m by +2-4 °C, which is similar to published data (231,233,234). The overall degree of stabilisation required for structural studies is protein dependent and moderate stabilisation might suffice if combined with other strategies such as nanobodies (235), fusion-proteins (58), truncations (236), high-affinity ligand binding (237), protein relipidation (238) or *in meso* crystallisation (61). Optimised GPCRs used in structural studies are usually stabilised by +10 °C or more (41). Based on this benchmark, the introduction of single variant G130A (ΔT_m= 13.0 ±1.4 °C) or F20Y (ΔT_m= 9.1 ±1.5 °C) may already be sufficient to facilitate structure solution. This is investigated in Chapter 5. Importantly, the experimental design did not aim to demonstrate statistical significance, which would require up to 12 experimental repeats (change: 2°C, 90% power, p-value = 0.05, average standard deviation: 2.3 °C), but identify biologically important changes that are beneficial for the biochemical or biophysical characterisation of target proteins.

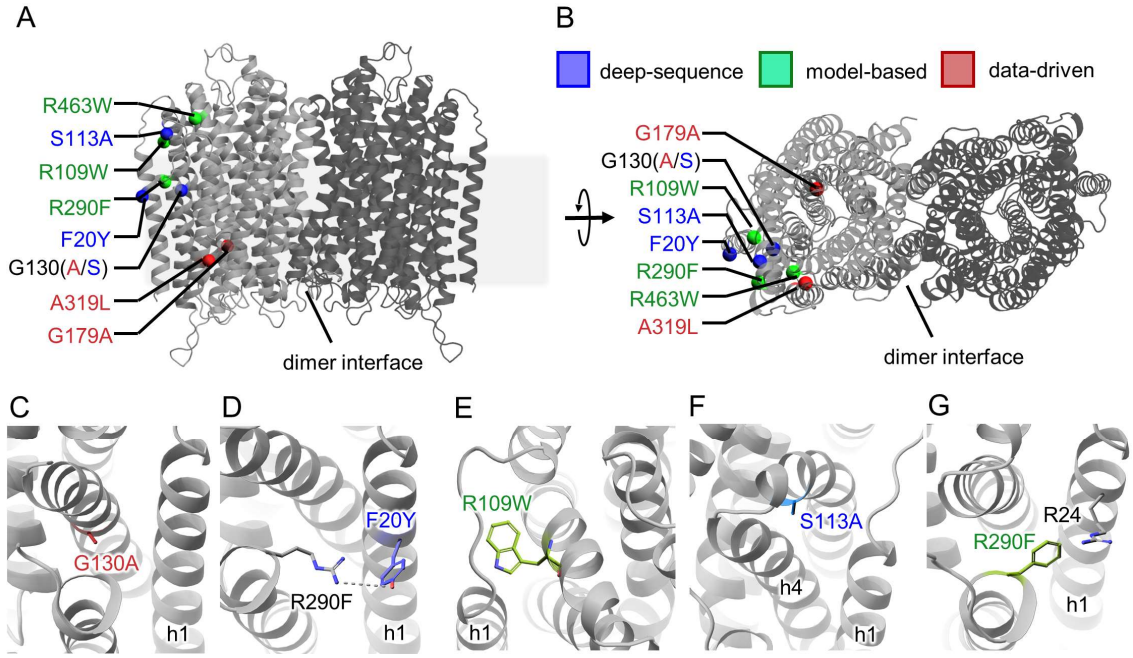


Figure 3.8: Top stabilising mutations of *Cl*-PPase mapped as coloured spheres to a comparative homology model. View from the membrane plane (A) and from the cytoplasmatic site (perpendicular to membrane plane) (B) showing the clustering of mutations opposite of the dimer interface. Rational of protein stabilisation with helix 1 (h1) and top stabilising mutations G130A (C), F20Y (D), R290F (E), R109W(F) and S113A (G) shown. Figure adapted from Harborne *et al.* (173).

Combining mutations can help to further stabilise challenging targets, but is not trivial as the effects of amino acid substitutions might not be additive if the underlying mechanism is based on the same principle (41). Indeed, the combination of top stabilisers G130A and F20Y did not further improve thermostability in *Cl*-PPase. Mapping all stabilising mutations to the comparative homology model revealed their clustering away from the protein dimer interface (Figure 3.8A-B). The underlying molecular mechanisms of protein stabilisation in this dynamic protein region (142) are probably based on improved helix-

helix interactions, rigidifying helix 1 in particular. This could explain why the combination of G130A and F20Y did not have additive power with respect to thermostabilisation. Both the removal of a helix breaking glycine by G130A (Figure 3.8C) and the potential for additional hydrogen bonding by F20Y substitution (Figure 3.8D), are capable of constraining helix 1 movement against helix 4 or 7, respectively. Similarly, R290F, R109W and S113A may confer increased protein stability by reducing electrostatic repulsion (*e.g.* to R24, Figure 3.8E) and improving hydrophobic packaging at the interface of helix 1 (Figure 3.8F-G). However, a robust analysis of stabilising mechanisms requires an experimentally determined high-resolution *Cl*-PPase structure (Chapter 5).

IMPROvER was strikingly successful at selecting active *Cl*-PPase variants that are suitable for functional characterisation and time-resolved biophysical studies, which was likely facilitated by filtering against a list of critical residues. However, locking a certain protein conformation can also be beneficial for structural studies. For example, stabilising GPCR variants are often selected based on bias towards a particular state by scanning mutagenesis in the presence of a high-affinity agonist or antagonist to aid crystallisation and resolve a particular conformation of interest (226,239). Such an approach is also compatible with IMPROvER and simply requires running and experimental testing with ligand-bound states.

Overall, IMPROvER and the established workflow proved very successful for the streamlined identification of stabilising mutations in *Cl*-PPase and other test proteins (173). This will enable more academic laboratories to stabilise challenging targets with reasonable effort and extend the scope of target size for stabilisation campaigns.

3.3.2. Development of a novel high-throughput lipid screen to stabilise membrane proteins

The RAMP lipid screen was specifically designed as a simple and cost-effective *multi-purpose* tool to aid the functional and biophysical characterisation of membrane proteins. Here, I focused on demonstrating its use as a quick, simple and cost-effective tool for protein stabilisation, which typically requires less lipid than dispensed in the screen (46,45). As a result, the solubilisation of lipid in detergent was incomplete but can likely be improved by sonication, more extensive manual mixing, addition of stirring bars or increasing the detergent concentration if compatible with downstream experiments. For stability screening, incomplete solubilisation and the final lipid concentration are irrelevant as long as sufficient amounts are solubilised as indicated by the various effects relipidation had on the protein stability of $A_{2A}R$, UapA and *Tm*-PPase (3.2.2.2).

Protein-lipid interactions, particularly the effect of CHL/CHS binding on stability, structure and function, have been studied in detail for GPCRs and served as a positive control here (34). For example, CHS was previously shown to improve $A_{2A}R$ stability and aided in its initial structure determination, which led to the identification of CHL binding sites (226,240). In order to purify sufficient amounts of $A_{2A}R$ in the absence of beneficial lipids and reinstall improved thermostability upon relipidation, a stable $A_{2A}R$ construct (Δ317-412, 209-218→BRIL) was used (240). Typically, stabilising effects conferred by CHS also translate to non-stabilised protein (241). Upon relipidation with CHS, thermostability improved in accordance with previously reported values (240). Similarly, antagonist binding (ZM241385) also rendered the protein more stable, although to a lesser extent than reported previously (226). These results provided confidence in our screening approach for stabilising lipids and stability measurements by nanoDSF. Slight discrepancies in the total degree of stabilisation are likely explained by differences in the assay conditions such as no CHS but high salt and theophylline being present. Additionally, the assay parameters (*e.g.* temperature ramp rate) also differed, which govern the T_m determination.

In $A_{2A}R$ screening, the next two best stabilising lipids were POPA and LMPC for which no specific protein-lipid interactions have been reported previously. PCs, including LMPC, might compete with CHL to modulate receptor activity as they associated with CHL binding sites in MD simulations (242) and have similar chemical properties (neutral, mostly hydrophobic). MD simulations further revealed the potential regulation of GPCR dynamics and conformational stabilisation by anionic phospholipids (33). PIP_2 was found to specifically bind more deeply and tightly to the intracellular side of the active receptor state (±mini-G_s) than to the inactive state (33). These lipid binding cavities might also be accessible to other phospholipids, particularly PAs such as POPA due to their small head groups, conferring conformational stabilisation.

Studies investigating protein-lipid interactions in UapA have mostly focussed on protein organisation and function. PEs, PIs and PCs were shown to co-purify in DDM and subsequent MD simulations and mutational studies revealed their role in stabilising the functional dimeric protein form (32). Here, PIs were too expensive to be included in the screen and PEs and PCs did not enhance resistance against thermal unfolding in UapA. Native lipids that remained bound to UapA during isolation may mask the effects of screen lipids if both compete for the same binding sites (32). However, both PSs included in the screen were found to confer thermostability in UapA, although not identified in previous lipidomic analysis. This indicates that their association is rather weak compared to PIs, PEs and PCs. The importance of PSs for dimer formation of the structurally related boron transporter from *S. cerevisiae* (ScBOR1p) suggests a similar role in UapA (214).

Stabilisation of UapA by CHS (and CHL) has also not been reported previously. UapA is physiologically exposed to ergosterol, the fungal equivalent to mammalian CHL. Due to its structural similarity, CHS/CHL might act as molecular mimic for ergosterol, thus conferring thermostabilisation (243). Surprisingly. UapA was not stabilised by xanthine as previously reported for protein solubilised in harsh detergents such as NG (244). Here, UapA was solubilised in DDM, a much milder detergent, which may already favour a more stable conformation. The implications of non-specific interactions between detergent micelles and proteins on stability are illustrated by and discussed for *Tm*-PPase in more detail.

Tm-PPase is a thermophilic protein and, as such, withstands much higher temperatures compared to the mesophilic UapA and $A_{2A}R$. Nevertheless, the average degree of stabilisation and number of stabilising lipids found was the highest out of all three test proteins. Moreover, stabilisation took place across almost all lipid classes. This suggests non-specific interactions of protein with mixed detergent-lipid micelles rather than specific protein-lipid interactions as cause of the broad stabilisation. Indeed, *Tm*-PPase was the only protein purified in OGNG, a rather harsh detergent routinely used in X-ray crystallography (129) due to its short alkyl chains, which allows tighter packing and better crystal contact formation at the cost of protein stability (245). The lipidation of short chain detergent micelles likely reduces curvature stress and leads to increased protein stability (246). This effect is less pronounced in more gentle, longer chain detergents such as DDM, which might unmask specific protein-lipid interactions in *Tm*-PPase but was reported to lead to a more complex thermal unfolding behaviour with two transitions (204). Only a few lipids did not stabilise *Tm*-PPase and changes in the lipid head group or tail within a lipid class were often sufficient to confer different effects on protein stability (Figure 3.7B). For example, LMPC and DMPC differ only in their hydrophobic tails, whereas anionic lipids POPA and POPS have a changed hydrophilic head group, highlighting that both the hydrophobic tail and the head group can influence stability (neutral *versus* stabilising). Overall, our understanding of lipid effects on M-PPase biology remains very limited. Soybean lecithin is known to enhance hydrolytic protein activity in thermophilic M-PPases but was not included in the screen due to its short shelf life. I also investigated binding of the non-hydrolysable substrate analogue IDP, which is known to protect against protein degradation and trigger conformational changes in *Tm*-PPase (129). The latter is reflected by a more pronounced change in intrinsic fluorescence upon unfolding as compared to the relipidated sample (Figure 3.7A). However, conformational stabilisation did not translate to thermal stabilisation, indicating that the IDP-bound state is energetically less favourable than the apo state under assay conditions.

MAGs were the only lipids that consistently conferred destabilising (or neutral) effects in all three test proteins. Although counterintuitive at first due to their important role as host lipid for LCP crystallisation, this is not surprising as MAGs were specifically designed to form highly curved mesophases (61,62). Therefore, they might induce curvature stress on and the disruption of detergent micelles under conditions that do not support mesophase formation (247). This is supported by the fact that protein solubilised in rather harsh detergent such as OGNG (*Tm*-PPase) was less affected by MAG addition than protein solubilised in gentle detergent such as DDM (A_{2A}R and UapA). In future versions of the screen, MAGs will be replaced by 10:0-18:0 PCs to extend the sampled lipid space and better assess the effect of different alkyl lengths and membrane curvatures on target proteins (248).

In summary, the application of the RAMP lipid screen in stabilisation campaigns of three test proteins demonstrated its potential as a tool for the cost-effective and rapid identification of stabilising lipids with minimal effort. Known stabilising effects were recapitulated using A_{2A}R as a control protein, thereby validating our approach. Moreover, novel lipid molecules interacting with A_{2A}R (POPA) and UapA (CHS, PSs) were identified, which provides a basis for further investigation. In this study, I used label-free stability screening by nanoDSF to evaluate the effect of lipid binding, but other assays such as fluorescence size exclusion chromatography (FSEC), circular dichroism (CD) spectroscopy or microscale thermophoresis (MST) could be employed too. However, these are less well suited for membrane-proteins. The screen could also be used as a source for lipids in other lipid-based techniques, for example in HiLiDe crystallisation by varying the lipid at a set detergent concentration. This highlights its value as an affordable and broadly applicable tool of membrane protein research.

3.4. Future directions

Both tools, IMPROvER and the RAMP lipid screen, proved successful for the streamlined and more rational stabilisation of membrane proteins and are already available to the scientific community upon request. In future, access will be simplified by running a webserver for IMPROvER and commercialisation of the RAMP lipid screen in cooperation with Molecular Dimensions.

Chapter 4. Structural and functional studies of a K^+-independent H^+-PPase

4.1. Introduction

Section 1.7 summarises how little we know about basic molecular processes in M-PPases. In particular, the role of the second conserved key change of K^+-(in)dependence (G/A$^{12.49}$T), ion pumping selectivity, substrate inhibition and functional asymmetry are all poorly understood. A more detailed mechanistic understanding of M-PPases requires structural data from all subclasses and functional studies of the key changes involved in K^+-(in)dependence. To date, only structures of a K^+-dependent Na^+-PPase and a K^+-dependent H^+-PPase have been published with no structures of different M-PPases released in ten years.

4.1.1. Summary of previous work on *Pa*-PPase

Former group members Dr. Juho Kellosalo and Dr. Craig Wilkinson worked towards the first structure of a K^+-independent M-PPase and established a protocol for the expression, purification and crystallisation of the naturally thermostable H^+-PPase from *P. aerophilum* (178,249). Initial vapour diffusion grown crystals were optimised by pH, salt, PEG and detergent screening as well as crystal dehydration to improve crystal diffraction (178). Dr. Wilkinson shot several hundred crystals at X-ray beamlines and combined the first few hundred images of the best diffracting crystals into one 3.8 Å dataset with improved quality (178). Nevertheless, side chain density was poor at the given resolution even when excluding data influenced by radiation damage. The data quality severely impaired the refinement process, particularly the manual modelling in coot, which was very laborious and difficult and could not be completed by Dr. Craig Wilkinson.

4.1.2. Aims and strategy

To address open questions of K^+-(in)dependence, ion pumping selectivity and functional asymmetry, I worked towards better quality crystals of the K^+-independent *Pa*-PPase by optimising vapour diffusion grown crystals or, alternatively, growing tightly packed type I crystals in LCP (1.1.3). Moreover, I continued working on the previously collected 3.8 Å dataset of *Pa*-PPase due to the SARS-CoV-2 pandemic limiting access to research laboratories and equipment. These efforts were supplemented by functional studies of wild-type and variant protein for a more complete picture of *Pa*-PPase biochemistry. I investigated the importance of key residues K$^{12.46}$ and T$^{12.49}$ at the cationic centre on K^+-dependence and substrate inhibition by assessing the hydrolytic activity of wild-type,

single ($K^{12.46}A$, $T^{12.49}A$) and double ($K^{12.46}A+T^{12.49}A$) variants at different K^+ and substrate concentrations.

4.2. Results

4.2.1. Heterologous protein expression and hot-solve purification

The expression and purification followed the protocol established by Dr. Juho Kellosalo (249) and Dr. Craig Wilkinson (178) and is described in 2.6.1 and 2.6.2. *Pa*-PPase was expressed in *S. cerevisiae* and purified *via* the "hot solve" method, which utilises elevated temperatures to enhance the detergent-extraction and heat precipitate mesophilic host proteins, thereby improving protein yield and purity (177). About 0.4 mg L^{-1} culture of wild-type and variant ($K^{12.46}A$, $T^{12.49}A$, $K^{12.46}A+T^{12.49}A$) *Pa*-PPase were obtained following this method (Figure 4.1), which was sufficient for their subsequent functional and structural characterisation. Wild-type protein was purified in DM and DDM, which yielded the best diffracting crystals in previous studies (178). Variant protein used for functional characterisation in quantitative P_i-release assays was purified in DDM only and supplied by group member Dr. James Hillier.

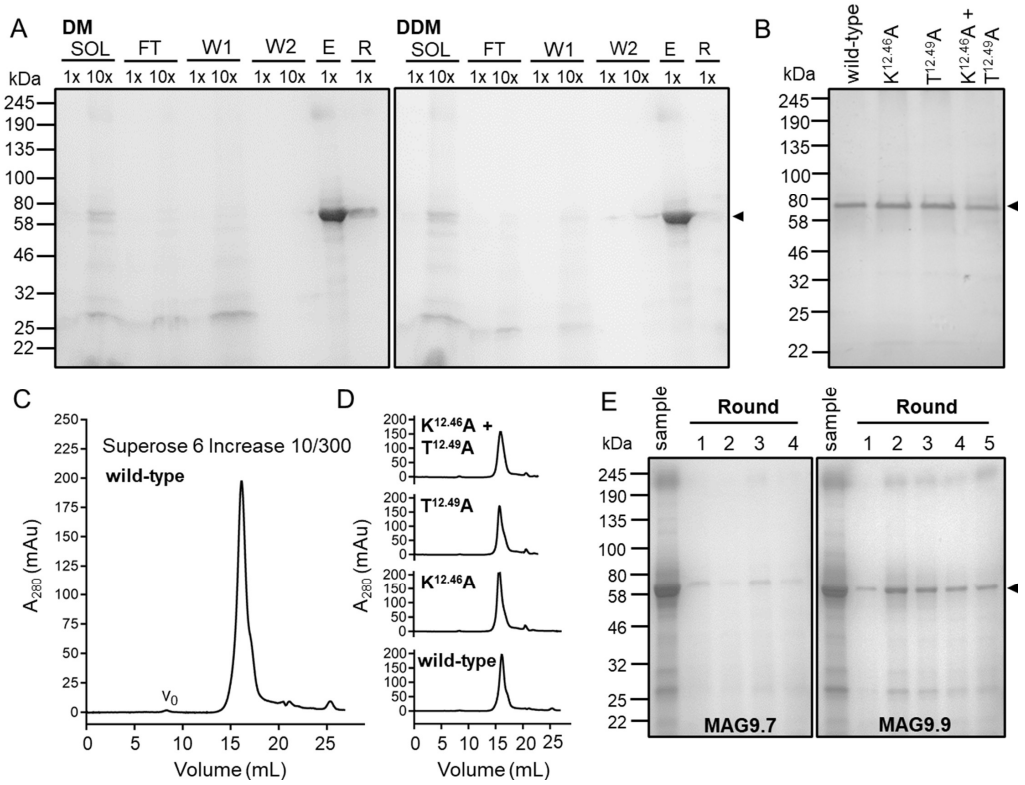


Figure 4.1: "Hot-solve" purification of *Pa*-PPase. (A) SDS-PAGE (Coomassie stain) analysis of wild-type *Pa*-PPase purified in DM and DDM. Following undiluted (1x) and concentrated (10x) samples were loaded: Detergent-solubilised crude membrane extract (SOL), IMAC flow-through (FT), column wash 1 (W1) and wash 2 (W2), protein elution (E) and IMAC resin after elution (R). (B) Purified wild-type and variant *Pa*-PPase in the absence of K^+ and Na^+ on Coomassie-stained SDS-PAGE. (C) Analytical SEC of wild-type *Pa*-PPase. The void volume is indicated by v_o. (D) SEC elution volume comparison of wild-type and variant *Pa*-PPase on Superose 6 Increase 10/300 column. (E) Analysis of protein reconstitution into mesophase during Cubicon. Purified wild-type *Pa*-PPase (DDM) and "protein-depleted" aqueous phase of individual Cubicon rounds loaded on SDS-PAGE (Coomassie stain). Black arrows indicate target protein bands running approximately at their expected size on SDS-PAGE gels.

There are only a few contaminating bands present in the IMAC elution fractions of wild-type (Figure 4.1A) and variant samples (Figure 4.1B) that were purified in DM and DDM. Most protein eluted from the resin in the presence of 50 mM imidazole. The SEC profile shows a monodisperse protein peak with a small shoulder towards larger elution volumes (Figure 4.1C). The elution volume and peak profile did not change upon mutagenesis, indicating that wild-type and variant protein have a similar hydrodynamic radius (Figure 4.1D). As thermostable M-PPases readily crystallise after IMAC elution following the "hot-solve" method and subsequent concentration (178), SEC was done for analytical purposes only. Unfortunately, wild-type *Pa*-PPase tended to precipitate above 10 mg mL^{-1} in both detergents used. To reach sufficient concentrations for LCP crystallisation, protein was enriched in the mesophase following the Cubicon method (1.1.3). The final concentration upon dilution with host lipid for LCP formation was estimated to be ~8 mg mL^{-1} (9.9MAG) and ~12 mg mL^{-1} (9.7MAG). However, due to incomplete transition of protein from the aqueous phase into the meso phase during reconstitution rounds, the actual concentration is slightly lower and not known (Figure 4.1E).

4.2.2. Vapour diffusion and lipidic cubic phase crystallisation

Vapour diffusion grown crystals that were obtained from conditions that previously yielded a 3.8 Å *Pa*-PPase dataset diffracted to 7 Å at best (Figure 4.2A), even after optimisation by finer pH, salt and PEG screening than done in the past (100 mM MES-NaOH pH 6.5, 29% v/v PEG400, 50 mM $LiSO_4$, 50 mM NaCl, 2 mM Na_4IDP). Similarly, crystals obtained from screening conditions that led to several *Tm*-PPase structures did not diffract better than 4.6 Å (100 mM Tris-HCl pH 8.0, 29% v/v PEG400, 10 mM KCl) (Figure 4.2B). The cell parameters of crystals from both conditions (a=106 Å, b=151 Å, c=260 Å, α=90 °, β=101°, γ=90°) also differed from those obtained by Dr. Craig Wilkinson (Table 4.1), which may explain the observed differences in diffraction quality.

Due to issues of reproducibility and limited success in improving crystal diffraction based on previous M-PPase crystallisation conditions, further optimisation focused on LCP crystallisation to obtain crystals diffracting beyond 3.8 Å. After extensive screening using several commercial sparce matrix screens and different host lipids (9.9MAG or 9.7MAG), microcrystals (2-4 µm) grew in 70 mM trisodium citrate pH 5.0, 70 mM NaCl and 21% *v/v* PEG500 DME (70% MemMeso, E3) in LCP with 9.7MAG as host lipid (Figure 4.2C). However, their diffraction was limited to ~20 Å at microfocus beamlines. Diffraction improved to ~7 Å after optimisation by pH, salt and PEG screening, which led to the formation of large (~20x100 µm) 2D plates (Figure 4.2C). These crystals were extremely fragile and fractured during harvesting, even when fishing the entire LCP bolus with a

mesh. Only a few diffraction images could be recorded in grid scans at microfocus beamlines, which was insufficient for indexing let alone structure solution. In subsequent crystallisation trials, LCP was dispensed into a range of *in situ* plates to simplify crystal access and remove the need for conventional harvesting. Out of the three systems tested, a prototype Swiss CI LCP crystallisation film, MiTeGen In situ-1 ™ plates and MiTeGen IMISX™ plates, crystals formed only IMISX™ plates, which are sealed with glass covers like the plates typically used. After crystal formation, the IMISX™ plates were sent to beamline i24 (Diamond) for *in situ* serial crystallography. Currently, a holder for mounting IMISX™ plates onto the beamline is being developed by beamline scientist Dr. Sam Horrell to test these crystals.

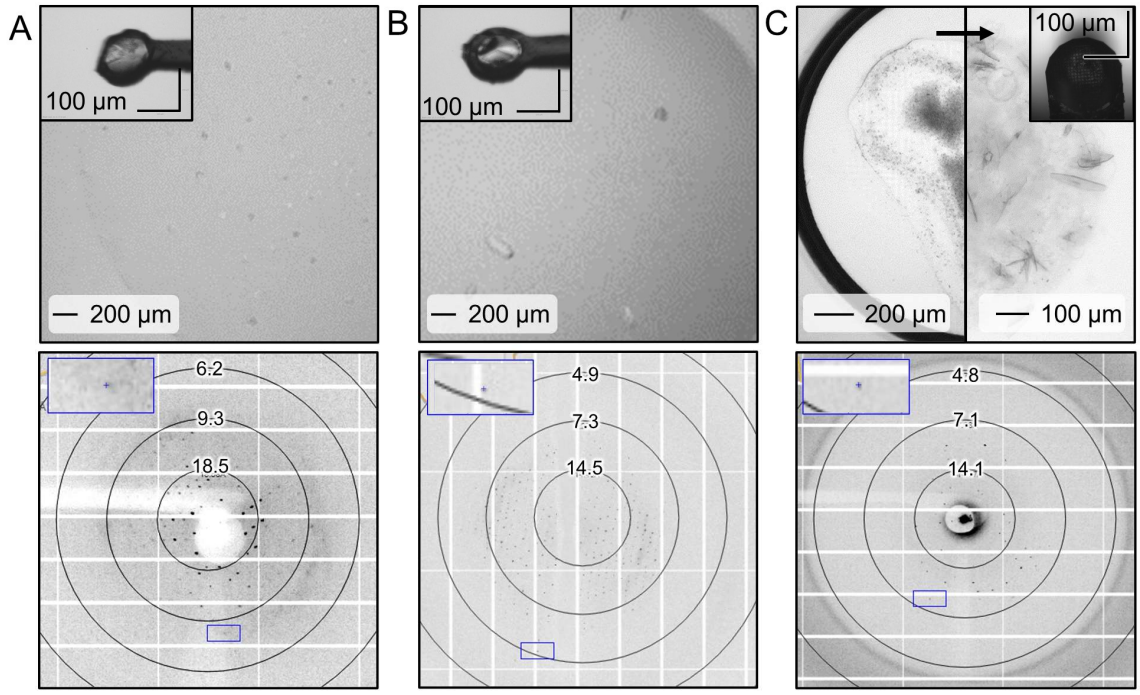


Figure 4.2: *Pa*-PPase crystallisation and data-collection. (A) Vapour diffusion grown crystals based on conditions identified by Dr. Craig Wilkinson. (B) Vapour diffusion grown crystals based on *Tm*-PPase crystallisation conditions. (C) LCP crystals obtained in commercial sparse-matrix screening (left panel) before and after optimisation (right panel). Upper left/right box shows harvested crystals in a loop. The bottom panels show the protein diffraction pattern with resolution rings from ~18 to 3 Å. Blue box highlights highest resolution diffraction spots.

Further crystal optimisation, for example by HiLiDe crystallisation or extended screening using short-chain detergents and phosphorus as well as non-phosphorus M-PPase inhibitors or a broader range of MAGs, was hampered by the SARS-CoV-2 pandemic limiting access to research laboratories. Therefore, I continued to work on the previously collected 3.8 Å dataset with the aim to compensate for the rather poor data quality as well as possible by extensive model refinement cycles and careful in-depth analysis of all possible rotamer orientations where side chain density was poor. In contrast to previous attempts at solving and refining this structure, the data was submitted to the StarAniso webserver for anisotropic truncation of merged intensity data, Bayesian estimation of structure amplitudes and anisotropic data correction prior to molecular

replacement. Afterwards, the structure was subject to many rounds (>300) of manual modelling and refinement (4.2.3).

4.2.3. Structure solution of a combined 3.8Å dataset

As summarised in 4.1.1, Dr. Craig Wilkinson had previously obtained a combined *Pa*-PPase dataset that he was unable to refine due to its low resolution. To improve the electron density map quality, I corrected the data for anisotropy using the StarAniso webserver (https://staraniso.globalphasing.org/) before solving the structure of *Pa*-PPase by molecular replacement (MR) (Table 4.1). The obtained resolutions for the three directions were 5.3 Å along the h axis, 4.1 Å along k and 3.8 Å along l (Table 4.1). I used a homology search model based on the *Tm*-PPase:Mg_5IDP structure (PDB: 5LZQ) with loops and hetero atoms removed for MR, which placed one *Pa*-PPase dimer in the asymmetric unit. The inspection of the newly obtained $2F_o$-F_c electron density maps showed that the transmembrane helix backbone could be reliably placed into the electron density at a contour level of 1 sigma (σ) (Figure S3, Appendix), allowing a good global overview and comparison to previously published M-PPase structures. However, loop densities were still absent except for short or highly ordered loops, for example $loop_{5\text{-}6}$ that folds over the active site upon substrate binding. There was also no electron density for regions rich in low molecular weight residues and residues with flexible side chains (Figure S5, Appendix), which includes key residues of K^+-(in)dependence ($K^{12.46}$) and ion-selectivity ($K^{16.50}$). The modelling of these side chains required a careful analysis of all possible rotamer conformations in context of the local environment and is justified in the text where necessary. Most side chains could be modelled with confidence when considering electrostatic interactions, steric hinderance and geometry restraints. Average NCS maps and feature-enhanced maps further helped with the modelling of poor electron density regions.

Positive density (F_o-F_c) was observed at 3 σ in both subunits of the *Pa*-PPase homodimer active site, at the ion gate and in the centre of the membrane-spanning section of the dimer interface (Figure 4.3A). I built a Mg_5IDP complex into the active site as seen in other IDP-bound M-PPase structures. Additionally, water molecules bridging between $loop_{5\text{-}6}$ and the metal cage were placed in regions with excess positive F_o-F_c density (Figure 4.3B). Another structural water was built into the positive F_o-F_c density at the ion gate, which also occurs in the high resolution structure of *Vr*-PPase:Mg_5IDP (Figure 4.3B). At the dimer interface I built a sulphate molecule (SO_4^{2-}) (Figure 4.3B), which fitted the local B-factor distribution and chemistry better than other ions present in the crystallisation conditions (Table 4.2). The electron density maps improved throughout the refinement process. The final R_{work}/R_{free} was 29.4%/31.3% with appropriate stereochemistry for the given resolution range (Table 4.1).

Table 4.1: X-ray data collection and refinement statistics of 3.8 Å *Pa*-PPase structure. Adapted from Wilkinson (178).

Data collection	
Space group	$P2_1$
Cell dimensions	
a, *b*, *c* (Å)	107.2, 88.0, 116.8
α, β, γ (°)	90.0, 106.9, 90.0
Source	DLS: i04/i24 ESRF: ID23-1/MASSIF1
Wavelength (Å)	0.9686
Resolution (Å)	19.97-3.84 (3.98-3.84)
Overall (Å)	3.8
along h axis	5.3
along k axis	4.1
along l axis	3.8
Measured reflections	65003 (2063)
Unique reflections	13069 (653)
Completeness (%)	87.6
$CC_{1/2}$	0.979
Mean I/σ(I)	3.9
Multiplicity	5
B-factors (Å^2)	112.11
R_{merge}	0.335 (1.524)
R_{meas}	0.375 (1.749)
R_{pim}	0.163 (0.827)
Refinement	
R_{work}(%)/R_{free}(%)	29.4/31.3
No. of atoms	9734
Protein	9694
Ligands	33
Water	7
No. of chains (ASU)	2
B-factors (Å^2)	102.66
Protein	102.69
Ligands/Ions	94.85
R. M. S. Deviations	
Bond lengths (Å)	0.002
Bond angle (°)	0.047
Ramachandran statistics (%)	
Favoured	96.61
Allowed	3.16
Outliers	0.23

Statistics for the highest-resolution shell are shown in parentheses

Table 4.2: Local B-factor distribution at dimer interface of 3.8 Å *Pa*-PPase structure

Modelling	**B-factor‡**	**Comment**
Residues*	74	Reference value
R636	81/75Ŧ	Positive F_o-F_c density after refinement
SO_4^{2-}	85	Good 2Fo-Fc density fit
PO_4^{3-}	83	Not present in conditions
Na^+	54	Positive F_o-F_c density after refinement
Mg^{2+}	56	Positive F_o-F_c density after refinement
Li^+	52	Positive F_o-F_c density after refinement
K^+	64	Not present in conditions
H_20	56	Positive F_o-F_c density after refinement
Cl^-	63	Positive F_o-F_c density after refinement
Ca^{2+}	67	Not present in conditions

‡ Isotropic B-factor model
* Average B-factor of residues within 5 Å of positive density
Ŧ Average B-factor of residue/ average B-factor of head group

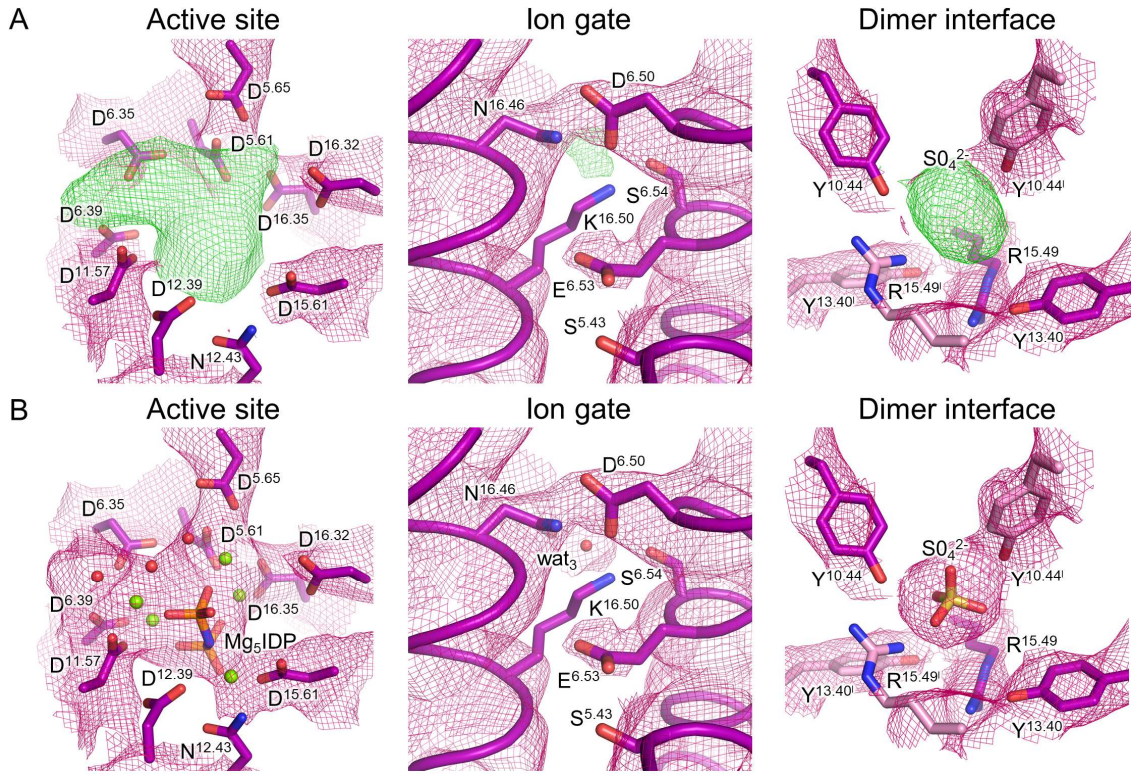


Figure 4.3: Electron density maps of *Pa*-PPase:Mg_5IDP key regions. (A) $2F_o$-F_c map of active site, ion gate and dimer interface residues and F_o-F_c omit map with positive density shown in green and red for ligand or heteroatom binding regions. (B) $2F_o$-F_c map of active site, ion gate and dimer interface residues with ligands and heteroatoms added to the model. Mg^{2+} are shown as green spheres and structural water molecules are shown as red spheres. Residues of subunit A are coloured in purple and residues of subunit B a coloured in pink (additionally marked with apostrophes).

4.2.3.1. Structural overview and comparison to previous structures

The solved structure is in an Mg_5IDP-bound state with $loop_{5\text{-}6}$ closed, SO_4^{2-} localised at the dimer interface and a structural water located at the ion gate (Figure 4.4). In what follows, structural alignments and rmsd calculations are based on the Cα atom of subunit A (both subunits are nearly identical; rmsd/C_α: 0.27 Å) unless stated otherwise. The overall structure is very similar to other published M-PPase structures with an average rmsd ($\overline{\mathrm{rmsd/C\alpha}}$) of 1.37 ±0.18 Å to IDP-bound structures, 1.41 ±0.15 Å to product-bound structures, and a rmsd/C_α of 1.67 Å to the resting state structure (Table S6, Appendix). In general, outer ring helices display more variability ($\overline{\mathrm{rmsd/C\alpha}}_{\mathrm{OR}}$: 2.01±0.67 Å) than inner-ring helices ($\overline{\mathrm{rmsd/C\alpha}}_{\mathrm{IR}}$: 1.27±0.32 Å) when compared to *Tm*/*Vr*-PPase:Mg_5IDP.

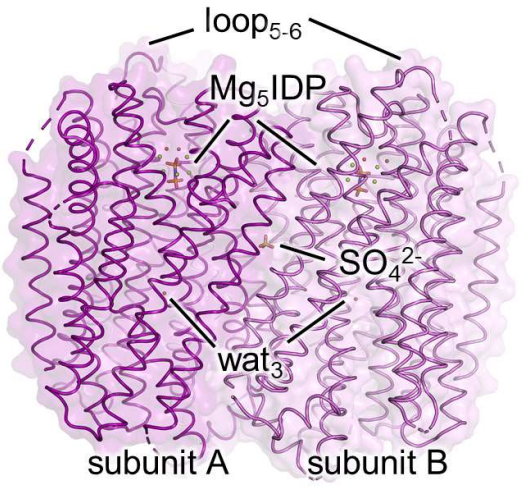


Figure 4.4: Structural overview of *Pa*-PPase:Mg_5IDP from the membrane plane.

Difference distances matrices (DiDiMa) allow the alignment-independent and straightforward identification of global conformational changes by comparing all inter-atom Cα distances between structures (188). This analysis highlighted outer ring helices 13-14 (and 2-3 in *Vr*-PPase but not *Tm*-PPase) as regions with major structural differences when comparing identical enzyme states (Figure 4.5A). In previously solved IDP-bound structures, helices 13-14 are bent halfway through the membrane by about 9° to remain near the cytoplasmic regions of helix 5 (Figure 4.5B). This enables propagation of motions from the inner to the outer ring and into the other subunit (indicated by apostrophe) *via* $E^{5.71}$–$R^{13.62}$–$R/I/K^{10.33'}$ (Figure S4A, Appendix). Until now, bending of helices 13-14 was always observed in IDP-bound structures and linked to substrate binding and subsequent $loop_{5\text{-}6}$ closure. In our new structure, the $E^{5.71}$–$R^{13.62}$–$R/I/K^{10.33'}$ interaction is lost (Figure S4B, Appendix) as cytoplasmic regions of helices 13-14 are straightened (Figure 4.5B) and further away from the enzymatic core, which resembles resting-state (Figure 4.5C) and product-bound (Figure 4.5D-E) structures more closely despite having IDP bound. This contradicts our previous model and suggests a different role for helix 13-14 movement in M-PPase function (4.3.5).

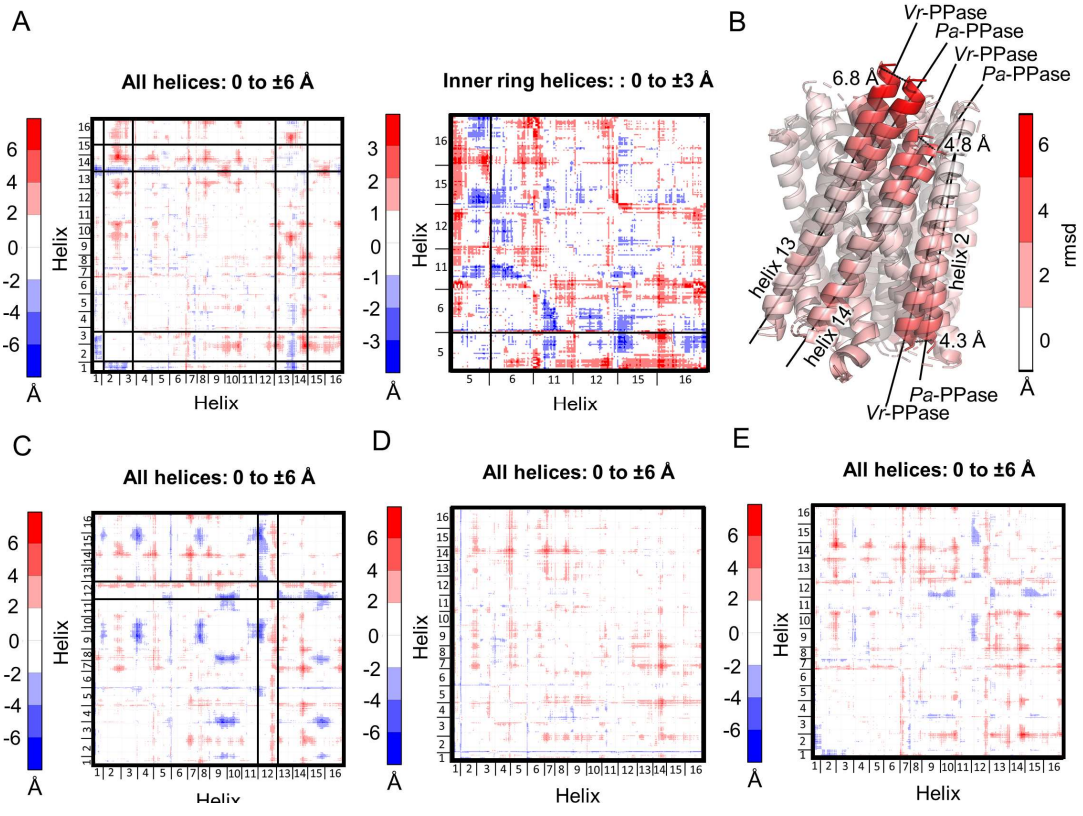


Figure 4.5: Structural comparison of *Pa*-PPase:Mg_5IDP to other M-PPase structures in difference distance matrices (DiDiMa). Inter atom distances shorter or longer in *Pa*-PPase than in other M-PPases are indicated by blue or red colour, respectively, in the matrices. Major structural changes are highlighted by black boxes. (A) DiDiMa of *Pa*-PPase:Mg_5IDP minus *Vr*-PPase:Mg_5IDP. Left panel shows the DiDiMa of all atoms, whereas the right panel shows inter-atom differences of inner ring helices only. (B)Structural alignment of M-PPase monomers with helices coloured by their rmsd/C_α. Dashed lines indicate the measured distances at terminal helix segments. (C) DiDiMa of *Pa*-PPase:Mg_5IDP minus *Tm*-PPase:CaMg. (D) DiDiMa of *Pa*-PPase:Mg_5IDP minus *Tm*-PPase:Mg_4P_{i2}. (E) DiDiMa of *Pa*-PPase:Mg_5IDP minus *Vr*-PPase:Mg_2P_i.

The changes of outer ring helixes cluster around helix 5 (Figure 4.6A-B), which is the only inner ring helix with notable above-average conformational differences to previously published IDP-bound structures ($\overline{\mathrm{rmsd/C\alpha}}_{\mathrm{h5}}$: 1.70±0.13 Å *versus* $\overline{\mathrm{rmsd/C\alpha}}_{\mathrm{IR}}$: 1.27±0.32 Å). Although the changes at helix 5 are less pronounced than for outer ring helixes, they are of particular interest to us due to the role of the inner ring in protein function. Outward movement (away from the inner ring) of helices 13-14 on the cytoplasmic site (Figure 4.6A) and helices 2-3 on the periplasmic site (Figure 4.6B) drag helix 5 along to maintain tight hydrophobic packaging, which effectively leads to its straightening (Figure 4.6C). These observations were confirmed by HELANAL-Plus, a helix geometry analysis software (191), which classified helix 5 as linear in *Pa*-PPase:Mg_5IDP but kinked in *Vr*-PPase:Mg_5IDP (Table S7). Additionally, helix 5 is more tightly wound in *Pa*-PPase:Mg_5IDP (Figure 4.5C) due to the presence of twice as many 3_{10} hydrogen bonds around $S^{5.43}$ and towards its flanking periplasmic segment (Table S8, Appendix). Consequently, the side chain orientations are different in this region as compared to *Vr/Tm*-PPase:Mg_5IDP, which is particularly interesting for $S^{5.43}$ as it is part of an enzymatic core region, the ion gate (4.3.5).

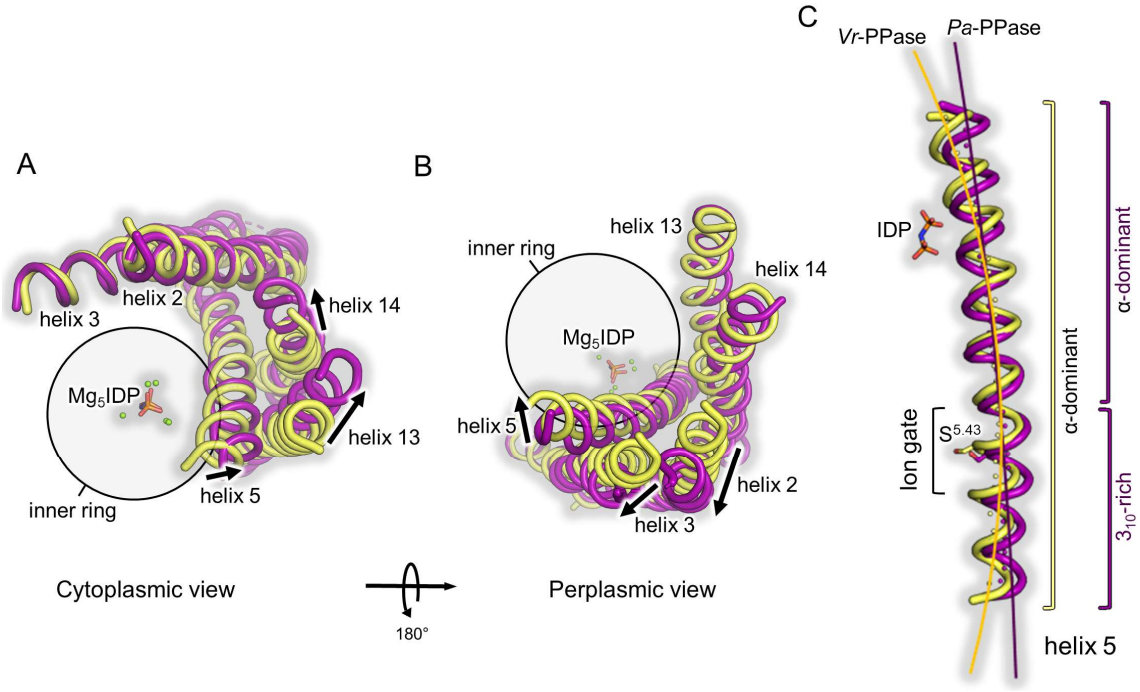


Figure 4.6: Comparison of helices 2-3, 5, and 13-14 orientations in *Pa*-PPase:Mg_5IDP to *Vr*-PPase:Mg_5IDP.. *Pa*-PPase:Mg_5IDP shown in purple and *Vr*-PPase:Mg_5IDP shown in yellow. Major conformational changes are indicated by arrows. (A) Outward movement of helix 13-14 on the cytoplasmic site, dragging helix 5 along. (B) Outward movement of helix 2-3 on the periplasmic site, dragging helix 5 along. (C) Close-up view of helix 5 showing its straightening in *Pa*-PPase:Mg_5IDP compared to *Vr*-PPase:Mg_5IDP. Helix straightening is highlighted by a curve running through the centre of helix 5 that was manually fitted to the local helix origin points, which are displayed as spheres in the helix centre.

4.2.3.2. Structural characteristics of K^+-independence

Most side chains involved in substrate coordination are well defined in the $2F_o$-F_c electron density map (Figure 4.3A, Figure S5 Appendix) if not stated otherwise in the text. This indicates good ordering and allows a reliable analysis of the active site conformation. The coordination of the Mg_5IDP complex by acidic residues appears to be almost identical to *Tm*/*Vr*-PPase:Mg_5IDP (rmsd/C_α: 0.81/0.73 Å, alignment of catalytic residues in the active site of subunit A), and the hydrolytic pocket volume is about 1200 $Å^3$ for all three structures (Figure 4.7A). However, I identified some interesting structural changes compared to previously solved K^+-dependent M-PPases with $A^{12.46}$ and $A/G^{12.49}$:

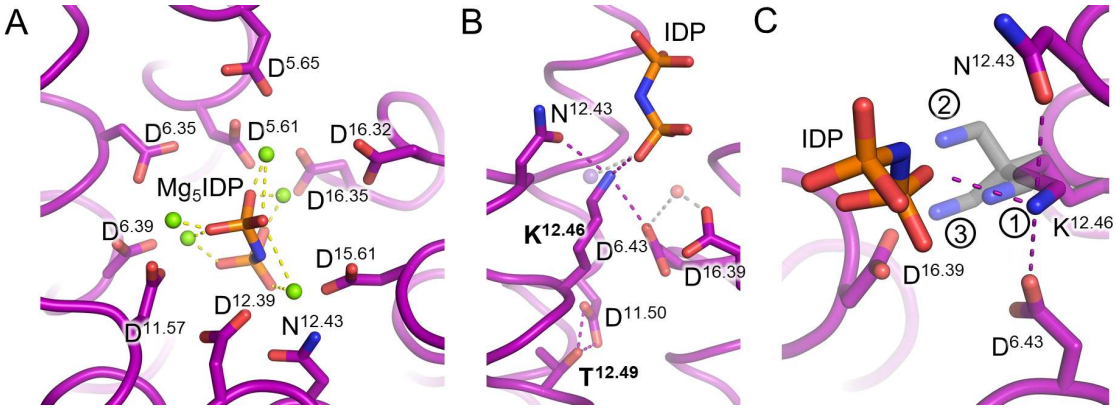


Figure 4.7: Structural overview of *Pa*-PPase active site with cationic centre. (A) Active site with IDP coordinated (dashed lines) in a Mg^{2+} metal cage (green spheres). (B) K^+/$K^{12.46}$ cationic centre with K^+ (transparent purple sphere) and nucleophilic water (transparent red sphere) modelled into the structure based on its position in *Vr*-PPase:Mg_5IDP. Key residues of K^+-independence are labelled in bold. Their interaction is shown by dashed lines. (C) Close-up view of $K^{12.46}$ with its three best rotamer conformations (grey).

First, the side chain of $K^{12.46}$ reaches into hydrolytic pocket of *Pa*-PPase. I modelled it to coordinate IDP and be held in place by $N^{12.43}$ and $D^{6.43}$. The ε-NH_3^+ group of $K^{12.46}$ is located at the K^+-binding site identified in *Vr*-PPase:Mg_5IDP (Figure 4.7B), thereby substituting for K^+ role in substrate coordination and explaining why *Pa*-PPase does not require K^+ for enzymatic activity. As stated earlier, there is mostly no electron density visible for lysine, including $K^{12.46}$ (Figure S5, Appendix), at the given resolution. Therefore, I evaluated all possible rotamer conformations for $K^{12.46}$ in terms of their hydrogen bonding potential and clashing with other residues (Table S9, Appendix) to model the most plausible pose. The built conformation of $K^{12.46}$ is based on rotamer 1 (Figure 4.7C) of the backbone-independent Richardson rotamer library used in Coot (250), which shows the smallest Van der Waals (VdW) radii overlap to nearby atoms (0.62 Å) and good hydrogen bonding ($n_{h\text{-}bonds}$: 3) (Table S9, Appendix). A clash is defined by a VdW radii overlap of ≥ 0.6 Å. Comparison to all other rotamer conformations clearly shows that this is the only conformation that avoids severe clashes. The two next best rotamers differ only in C_3 and can also explain K^+-independence as they remain near the K^+-binding site and coordinate IDP (Figure 4.7C). Therefore, this data provides

first structural evidence for $K^{12.46}$ as sole residue required for K^+-independence in M-PPases.

Second, $T^{12.49}$ interacts with $D^{11.50}$ (Figure 4.7B) in *Pa*-PPase, which is a direct result of the coupled A/G$^{12.49}$T change in K^+-independent M-PPases. I later investigate its putative involvement in substrate inhibition (4.2.4), as this control mechanism is lost in single A$^{12.46}$K variants (136) and $T^{12.49}$ has not been ascribed a functional role yet despite being evolutionary conserved in K^+-independent M-PPases.

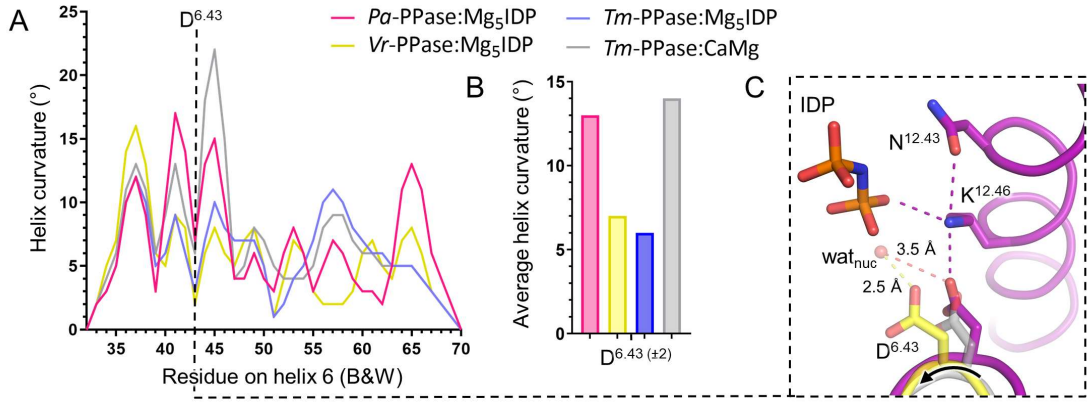


Figure 4.8: Local helix curvature analysis of M-PPase structures. (A) Residue by residue Bendix analysis of helix 6. Position 6.43 highlighted by a dashed line. (B) Average helix curvature around $D^{6.43}$ (±2 residues). (C) $D^{6.43}$ anti-clockwise rotation towards IDP for nucleophile activation. Coordination of key residues and wat_{nuc} shown as dashed lines. *Pa*-PPase:Mg_5IDP is shown in purple, *Tm*-PPase:CaMg in grey *Vr*-PPase:Mg_5IDP in yellow with wat_{nuc} displayed as transparent red sphere. The $D^{6.43}$ rotation is indicated by black arrow.

Third, the average helix curvature around $D^{6.43}$ is high (13°) compared to IDP-bound structures (6-7° in *Tm*/*Vr*-PPase:Mg_5IDP) and resembles that of a resting-state conformation (14° in *Tm*-PPase:CaMg) although the peak curvature drops from 22° to 17° (Figure 4.8A-B). As a result, the characteristic α-x-π-π-α hydrogen bonding pattern of the IDP-bound states is shifted by one residue in this region (Table S8, Appendix). Moreover, the incomplete release of strain energy stored in helix 6 in *Pa*-PPase leads to an incomplete anti-clockwise rotation of $D^{6.43}$ into the active site. This conformation is stabilised by a salt bridge that forms between $D^{6.43}$ and $K^{12.46}$. $D^{6.43}$ is still able to activate the nucelophilic water (wat_{nuc}), but its activation potential is reduced as the distance to wat_{nuc} (modelled as seen in *Vr*-PPase:Mg_5IDP) increases from 2.5 Å to 3.5 Å (Figure 4.8C).

$D^{6.43}$ is one of the few highly conserved residues at the active site that were only poorly defined in the $2F_o$-F_c density map. However, I am confident in the built conformation as $D^{6.43}$ was initially modelled as in *Tm*/*Vr*-PPase:Mg_5IDP, but this led to the appearance of positive and negative F_o-F_c density that guided its rearrangement as described above (Figure S5B, Appendix). Moreover, the new pose does not clash, is stabilised by a salt bridge and supported by additional side chain density that becomes visible at ~0.8 σ in the $2F_o$-F_c map (Figure S5C, Appendix).

4.2.3.3. Insights into ion-selectivity in K^+-independent H^+-PPases

The $2F_o$-F_c electron density map quality at the ion gate of subunit A is good for the given resolution and all residues except for $K^{16.50}$ are defined reasonably well, including the semi-conserved glutamate (Figure 4.3B). Side-chain orientations in subunit B, where the electron density maps were much poorer, were modelled to mirror subunit A. This is in line with the idea that binding of IDP to both subunits imposes symmetry on M-PPases as seen in all high-resolution M-PPases structures (1.5.2). I evaluated the different rotamer orientations of the backbone-independent Richardson library used in Coot for $K^{16.50}$ to reliably place its side chain. The chosen rotamer does not clash and has the best hydrogen bonding ($n_{h\text{-}bond}$ =4), making it superior to all other rotamer options (Table S10, Appendix). Although neither of the next two best rotamer orientations clash, I built the $K^{16.50}$ conformation that most closely resembles its position in *Vr*-PPase:Mg_5IDP as this provides much better coordination and places the ε-NH^{3+} group at the Na^+-binding site in *Tm*-PPase (Figure 4.9A), which fits the exclusively H^+-pumping nature of *Pa*-PPase.

All other side chain orientations and interactions at the ion gate of *Pa*-PPase:Mg_5IDP resemble the *Vr*-PPase:Mg_5IDP structure, despite the shift of the semi-conserved glutamate ($E^{6.53}$ in *Pa*-PPase, $E^{6.57}$ in *Vr*-PPase) (Figure 4.9B). This observation does not fit our current model of ion selectivity but is the logical consequence of *Pa*-PPase biology as *both* proteins are H^+-PPases. It also leads to the question of what other structural changes at the ion gate can explain the reorientation of $E^{6.53}$ in *Pa*-PPase?

Based on the results obtained from global structural comparisons in 4.2.3.1, the ion gate region around helix 5 was inspected for structural changes explaining $E^{6.53}$ reorientation. Indeed, helix 5 moved out of the protein core by about 2 Å as compared to other IDP-bound states due to its straightening and tighter winding in *Pa*-PPase:Mg_5IDP (Figure 4.9C-D). As a result, $S^{5.43}$ has sufficient space to adapt a conformation that stabilises $E^{6.53}$ facing away from the Na^+-binding site (Figure 4.9C) as seen in *Vr*-PPase:Mg_5IDP (Figure 4.9D). $E^{6.53}$ is likely protonated at this stage to neutralise the negative charge in the centre of the membrane-spanning protein section. In previous IDP-bound structures, helix 5 is closer to helix 6 and helix 16, forcing the semi-conserved glutamate to point away from $S^{5.43}$ (Figure 4.9C), thereby contributing to the formation of a Na^+-binding site (*Tm*-PPases:Mg_5IDP), or requiring its repositioning one helix turn down (*Vr*-PPases:Mg_5IDP) (Figure 4.9D).

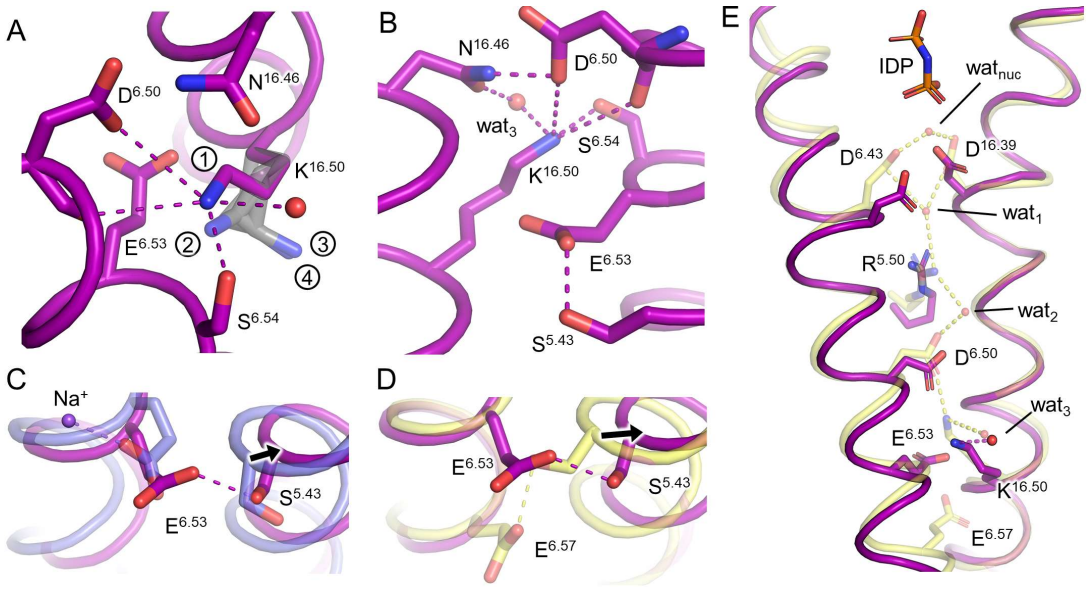


Figure 4.9: Structural overview of the ion gate and coupling funnel in *Pa*-PPase:Mg_5IDP. Dashed lines shown the coordination of key residues involved in ion selectivity such as $S^{5.43}$, $E^{6.53}$ and $K^{16.50}$. Structural water (wat_3) is displayed as a red sphere. Major structural changes are indicated by black arrows. (A) Close-up view of the *Pa*-PPase:Mg_5IDP ion gate structure. (B) Ion gate configuration with the four best rotamer conformations for $K^{16.50}$ shown in grey. (C) Comparison of the semi-conserved glutamate ($E^{6.53/57}$) orientation and helix 5 conformation in *Pa*-PPase:Mg_5IDP (purple) and *Vr*-PPase:Mg_5IDP (yellow). (D) Comparison of the semi-conserved glutamate ($E^{6.53/57}$) orientation and helix 5 conformation in *Pa*-PPase:Mg_5IDP (purple) and *Tm*-PPase:Mg_5IDP (blue). (E) Comparison of the residue orientation at the coupling funnel in *Pa*-PPase:Mg_5IDP (purple) and *Vr*-PPase:Mg_5IDP (yellow, transparent). The H^+ translocation pathway is shown for *Vr*-PPase:Mg_5IDP with water displayed as red transparent spheres and hydrogen bonds represented by dashed lines. The structural water (wat_3) in *Pa*-PPase:Mg_5IDP is displayed as non-transparent red sphere.

The reported changes at helices 2-3 and 13-14 (4.2.3.1) are likely linked to the orientation of helix 5 due to their proximity (Figure 4.6A-B) and direct interaction with it, for example *via* $E^{5.71}$-$R^{13.62}$ (Figure S4B, Appendix). Therefore, changes to the bend angle or winding of helix 5, *i.e.* the main chain hydrogen bonding pattern, may be propagated through helix 13 into the dimer interface and *vice versa*, linking inter-subunit communication to conformational changes defining ion selectivity. Taken together, these findings suggest that the orientation of helix 5 plays an important role in ion selectivity, which is not considered in the current model (4.3.5).

Lastly, I built a structural water molecule (wat_3) into positive F_o-F_c difference density at the ion gate (Figure 4.3B). This has been seen before in high-resolution H^+-PPase structures and is part of the conserved proton translocation pathway ($D^{6.43}$/$D^{16.39}$-wat_1-$R^{5.50}$-wat_2-$D^{6.50}$-$K^{16.50}$-wat_3) (Figure 4.9E).

4.2.3.4. Inter-subunit communication

The dimer interface of *Pa*-PPase is formed by helices 10, 13 and 15 and somewhat different to other M-PPases (Figure 4.10A). Usually, non-polar amino acids are conserved at position 10.44 (97.6% conserved) and 15.49 (95.2% conserved). In *Pa*-PPase these positions are occupied by a tyrosine and arginine,

respectively (Figure 4.10B). The additional hydrogen-bonding potential and positive charge leads to the formation of an anion binding site, modelled as occupied by SO_4^{2-} (4.2.3), which has not been observed before. SO_4^{2-} mediates the inter-subunit communication of $Y^{10.44}$, $Y^{13.44}$, and $R^{15.40}$ (Figure 4.10A) in *Pa*-PPase:Mg_5IDP, whereas $Y^{13.44}$ directly interacts with its counterpart in the opposing subunit ($Y^{13.44'}$) in *Tm/Vr*-PPase:Mg_5IDP.

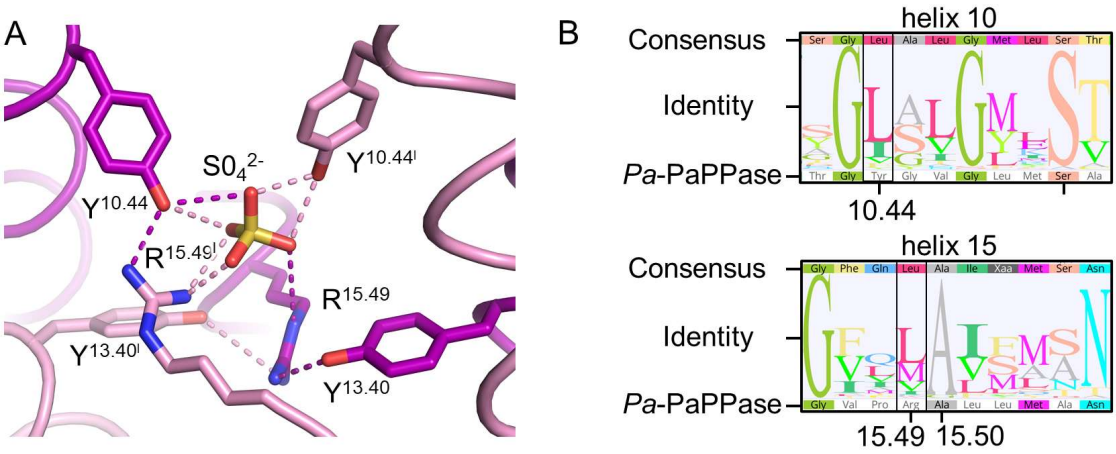


Figure 4.10: SO_4^{2-} binding site at the dimer interface of *Pa*-PPase:Mg_5IDP. (A) Structural overview with subunit A in purple and subunit B in pink (additionally marked with apostrophes). Side chain interactions are shown as dashed lines. (B) Sequence analysis of the SO_4^{2-} binding site. The consensus sequence and sequence identity (sequence logo showing the graphical representation of the residue conservation) are based on an alignment of 45 homologous sequences to *Pa*-PPase identified in a blastp search of the UniProt database. Residues of interest are highlighted by a black box and labelled following the B&W convention.

4.2.4. Functional studies of K^+-dependence and substrate inhibition

The unaccounted role of the evolutionary conserved A/$G^{12.49}$T change prompted me to investigate its possible involvement in substrate inhibition as mechanistic details are unclear and functional data from our group (131) and others (136) showed the loss of this regulatory effect when the two evolutionary key changes are uncoupled in $A^{12.46}$K variants (1.5.2). This hints towards a possible role of $T^{12.49}$, which interacts with $D^{11.50}$ in the *Pa*-PPase:Mg_5IDP structure, in substrate inhibition and functional asymmetry. Furthermore, I investigated the importance of $K^{12.46}$ for K^+-independence in *Pa*-PPase functional studies to confirm previous reports (136). The effect of the reverse mutations $K^{12.46}$A, $T^{12.49}$A and both $K^{12.46}$A+$T^{12.49}$A on substrate inhibition and K^+-(in)dependence was investigated in the hydrolytic activity assays of wild-type and variant protein under varying conditions using a quantitative colorimetric P_i-release assay (Figure 4.11).

All *Pa*-PPase samples were purified in the absence of K^+ *and* Na^+, as Na^+ is known to substitute for K^+ role in substrate coordination (136). Initially, parameters such as the protein concentration and incubation time at its temperature optimum (80 °C) were adjusted for *Pa*-PPase activity measurements (Figure 4.11A). These parameters were chosen to ensure P_i-release is linear with time, which allows the extraction of initial reaction velocities required for subsequent kinetic analysis by fitting the data to our

current best model of M-PPase catalysis (Figure 4.11B). The P_i-release was linear up to a 6-minute incubation and a protein concentration of about ~2.5 µg mL^{-1} (Figure 4.11A). At higher protein concentrations a shorter incubation time is required, but this is not compatible with the experimental procedure (2.7.3).

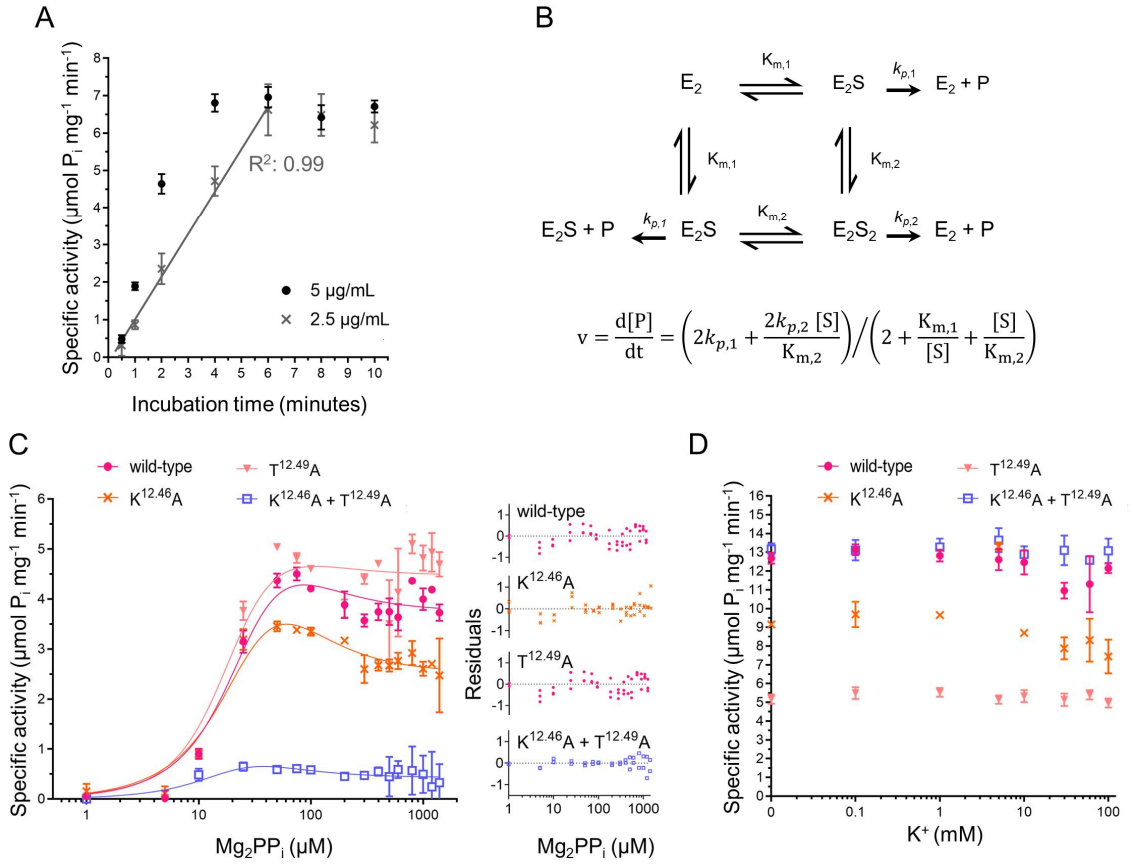


Figure 4.11: *Pa*-PPase quantitative P_i-release activity assays. (A) Assay optimisation to ensure P_i release by wild-type *Pa*-PPase is linear with time. P_i release at different protein concentrations and incubation times at temperature optimum (80 °C). Data fitted by linear regression. (B) „half-of-the-sites“ reactivity model and corresponding rate equation adapted from Artukka and co-workers (136). E_2: dimeric enzyme; E_2S and E_2S_2: enzyme-substrate complexes; P: product; K_m: Michaelis Menten constant; k_p: reaction rates of enzyme-substrate complexes. (C) Substrate-inhibition curves for wild-type and variant *Pa*-PPase recorded at 80 °C in presence of 100 mM K^+ and 1-1400 µM Mg_2PP_i (left panel). The data was fitted using the „half-of-the-sites“ reactivity model proposed by Artukka and co-workers (136). The residual plot shows a non-random wave-like pattern, indicating a poor fit (right panel). (D) K^+-(in)dependence assays using wild-type and variant *Pa*-PPase. The hydrolytic activity was measured at 75 µM Mg_2PP_i and varying K^+ concentrations (0-100 mM) at 80 °C.

Unfortunately, using high-temperatures caused issues during the P_i release assays that were designed to investigate substrate inhibition and K^+-(in)dependence. Mg_2PP_i showed a tendency to precipitate in the reaction condition, particularly at high substrate concentrations. This reflects in large error bars (Figure 4.11C). The incubation time and temperature could not be lowered without dropping below the assay sensitivity limit to detect P_i-release by *Pa*-PPase variants, nor could the enzyme concentration be increased without violating the $[S]>>[E_{total}]$ assumption (substrate in large excess to enzyme, so $[S_{t=0}] = [S]$) required for a simplified derivation of rate equations. Moreover, the use of pyrophosphoric acid as a Na^+/K^+-free source of PP_i also led to a high

background reading due to auto-hydrolysis, which reduces the sensitivity and assay range of P_i detection. Both issues contributed to poor reproducibility. For example, the double *Pa*-PPase variant $K^{12.46}A+T^{12.49}A$ was least active in functional assays with varying substrate concentrations (Figure 4.11C) but most active in functional assays with varying K^+ concentrations (Figure 4.11D). There was a 25x difference in specific activity comparing identical conditions (75 µM Mg_2PP_i, 100 mM K^+) of these assays.

Consequently, the substrate inhibition data did not fit the „half-of-the-sites" reactivity model proposed by Artukka and co-workers as indicated by poor R^2 values (0.83-0.92) and a wave-like pattern in residual plots (Figure 4.11C). The data also contradicts previous studies as all tested *Pa*-PPase variants showed substrate inhibition (Figure 4.11C) and K^+-independence (Figure 4.11D). Taken together, these issues draw the data reliability into question.

4.3. Discussion

Despite the availability of M-PPase structures in different catalytic states, the limited sub-group diversity left many mechanistic key questions of M-PPase biochemistry unanswered, for example: What is the role of the $A/G^{12.49}K$ coupled change in K^+-independent M-PPases; how is the $K^+/K^{12.46}$ cationic centre linked to substrate inhibition and functional asymmetry; and what defines ion selectivity in K^+-independent H^+-PPases and Na^+,H^+-PPases? The presented K^+-independent H^+-PPase structure from *P. aerophilum* addressed some of these questions and contributed to better understanding of M-PPase biochemistry and the ion-selectivity mechanism in particular.

4.3.1. Efforts towards a high-resolution *Pa*-PPase structure

My initial efforts towards a high-resolution *Pa*-PPase structure had only limited success. Vapour diffusion grown crystals obtained by Dr. Craig Wilkinson could not be reproduced (different cell dimensions) nor could the diffraction of crystals obtained in this project be improved beyond ~4.6 Å (178). Although following the same expression and purification protocol, I found that different protein batches influence crystal growth and diffraction quality quite significantly, leading to issues of reproducibility. This might be caused by differences in final detergent concentration due to the accumulation of protein-free micelles alongside protein-containing micelles during concentration using centrifugal filters (251,252). The degree of detergent concentration does not only variy between centrifugal filter suppliers but also within units from the same supplier (up to 20% difference), which leads to a unknown and variable final detergent concentration in each purification batch (253). Moreover, if different detergents are used throughout the purification, in this study DDM for solubilisation and DM for column elution, the final ratio of detergents in the sample might differ between protein batches due to an incomplete exchange. Ideally, the detergent exchange should go from a high CMC detergent to a

low CMC detergent as sufficient washing/dilution leads to the disruption of micelles formed by the detergent to be replaced. Due to the success of DDM in protein solubilisation (12) and its low CMC (~0.0087% *w/v*) this is not always possible (254).

Better diffraction can probably be obtained simply by testing many more *Pa*-PPase crystals at beamlines. Previously, only a few vapour diffusion grown crystals diffracted to up to 3.8°Å out of hundreds tested (178). The diffraction of most crystals was limited to 4-6 Å, similar to those obtained in this study. However, such a brute-force approach is cost and labour intensive and does not guarantee a dataset good enough to address open questions about the molecular mechanism of M-PPase biochemistry, which requires a significant improvement in diffraction quality (<3 Å). Other strategies such as conformational stabilisation (234), purification in shorter-chain detergents (255) or LCP crystallisation (62) follow a more rational approach, but were cut short by the SARS-CoV-2 pandemic limiting access to laboratories. Nevertheless, I was able to make initial progress towards a high-resolution *Pa*-PPase structure by obtaining the first ever M-PPase crystals *in meso*. Upon optimisation of initial hits, the crystal diffraction improved significantly but did not extend beyond ~7 Å. LCP crystals were extremely fragile and fractured during harvesting, which likely reduced the diffraction quality and hampered structure solution. In subsequent crystallisation trials using *in situ* plates, crystal formation was observed in the MiTeGen IMISX™ system, which is the only *in situ* plate using glass covers that are removed prior to X-ray diffraction experiments (256). The transition into glass-free systems requires further fine tuning, for example lowering the precipitant concentration to counter the enhanced evaporation. Unfortunately, IMISX™ plates are not yet compatible with the microfocus macromolecular crystallography beamline (i24) or the versatile macromolecular crystallography *in situ* beamline (VMXi) at Diamond. Therefore, a few *in situ* plates with and without crystals were sent to Dr. Sam Horrell who will 3D print a plate holder compatible with these beamlines. I continued to work on the previously collected 3.8 Å dataset in a complementary approach, which yielded the first publication-quality structure of a K^+-independent H^+-PPase.

4.3.2. Reliability of the structure interpretation at 3.8 Å

The electron density maps into which a *Pa*-PPase:Mg_5IDP model was built were good for the given resolution, which may be a result of merging non radiation-damaged parts from several data sets, anisotropy correction and careful iterative refinement. I was able to build most side chains of the enzymatic core regions as they are well ordered in presence of the non-hydrolysable PP_i analogues such as IDP. Additionally, feature-enhanced maps, NCS averaged maps and high-resolution reference structures were used to guide the modelling of poorly defined residues. I mirrored side chain

conformations from high-resolution structures in the same catalytic state for regions of poor electron density given the high sequence and structure conservation across M-PPases. For mechanistically important residues, I explored all rotamer options and gave detailed reasoning for my pick (4.2.3.2, 4.2.3.3) to justify the structural interpretation.

For example, $K^{16.50}$ at the ion gate is the only residue without sufficient electron density in this region but the orientation of nearby residues restricts the orientation of $K^{16.50}$ itself, allowing me to place it in the most stereochemically plausible pose. The chosen rotamer does not clash, shows good hydrogen bonding and places ε-NH^{3+} at a Na^+-binding site as seen in *Vr*-PPase:Mg_5IDP and in agreement with *Pa*-PPase biology as an exclusive H^+-pump. Importantly, the updated model of ion selectivity in M-PPases does not rely on the position of $K^{16.50}$ but is based on the orientation of helix 5, which is well defined in the electron density even at lower resolution, and $E^{6.53}$, also well defined in the electron density.

4.3.3. Structural basis for K^+-independence

Overall, the *Pa*-PPase:Mg_5:IDP structure is very similar to that of known M-PPases, which is not surprising given that sequence alignments identified only two conserved changes linked to the evolution of K^+-independence (141) and the high sequence identity amongst M-PPases (average pairwise identity: 41.8%). Significant structural changes are mostly limited to outer ring helices except for helix 5 and a few local changes of mechanistic key residues. Nonetheless, these changes allow me to explain the structural basis of K^+-independence and give insight into the mechanistic details of ion selectivity and functional asymmetry. The obtained structure supports mutational studies identifying $K^{12.46}$ as the sole determinant defining the requirement of K^+ for maximal enzymatic activity (136) by functionally and structurally replacing K^+ role in substrate coordination. K^+-independence is therefore the result of a few local changes rather than global rearrangements. $K^{12.46}$ also stabilises an energetically less favourable conformation of $D^{6.43}$ (high helix curvature), hampering the full clockwise rotation of $D^{6.43}$ into the active site and weakening the activation of wat_{nuc}. Why does this not abolish the activity of the enzyme? I posit two answers. The first one is that water becomes easier to deprotonate ($pK_{a,\ 20\ °C}$: 13.99 → $pK_{a,\ 100\ °C}$: 12.25 (257)) at the physiological temperature (~100 °C) of *Pa*-PPase (258), possibly allowing a nucleophilic attack on PP_i without full activation of wat_{nuc} by $D^{6.43}$. The second one is that the direct coordination of substrate by $K^{12.46}$ facilitates the charge transfer onto PP_i due to its ability to donate a full proton, unlike K^+, which can only provide electrostatic stabilisation. In summary, the evolution of K^+-independent M-PPases appears to only require the replacement of a single residue

at the active site, which is consistent with biochemical data showing K^+-independence for several M-PPases with a single $A^{12.46}K$ mutation (136).

Nonetheless, two coupled changes occur in K^+-independent M-PPases: $A^{12.46}K$ and $G/A^{12.49}T$. The unaccounted role of the evolutionary conserved $A/G^{12.49}T$ change prompted me to investigate its possible involvement in substrate inhibition (4.2.4) as mechanistic details are unclear and functional data from us and others showed the loss of this regulatory effect when the two evolutionary key changes are uncoupled in $A^{12.46}K$ variants (136). However, these studies were hampered by a series of issues using the quantitative P_i release assays at elevated temperatures that are discussed later (4.3.6).

4.3.4. Physiologically relevance of SO_4^{2-} binding at the dimer interface

The anion binding site at the dimer interface was modelled to be occupied by SO_4^{2-} based on the $2F_o$-F_c and F_o-F_c electron density maps, local chemistry and the B-factor distribution of nearby residues. However, there are no reports of SO_4^{2-} mediating *Pa*-PPase (or M-PPase) function and the protein was shown to be active in its absence, which questions its physiological relevance. Indeed, SO_4^{2-} cannot be distinguished from PO_4^{3-} at the given resolution. Although PO_4^{3-} is physiologically relevant for M-PPases, it was not modelled due to its absence in the crystallisation condition. Consequently, it cannot be ruled out that PO_4^{3-} was carried over from purification and is in fact bound in the *Pa*-PPase crystal structure.

Structural differences at the dimer interface, such as the binding of an anion, are particularly interesting as they could give mechanistic insights into intra-subunit communication in M-PPases. However, the different residue identities at positions 10.44 and 15.49 that lead to the formation of an anion binding site in *Pa*-PPase are not conserved in any M-PPase subclass with distinct biological characteristics, nor were they found in any other M-PPase sequence gathered in a blastp search of the uniport database. Consequently, anion binding to the dimer interface is likely *Pa*-PPase specific. It neutralises the positive charge of both $R^{15.49}$ head groups in a hydrophobic environment and may therefore by functionally and structurally important in *Pa*-PPase, but not for other M-PPases.

4.3.5. An updated model of ion-selectivity in M-PPases

The current model of ion selectivity in M-PPases is indicative of our limited in-depth insight into M-PPase biochemistry. It focusses on the position of a semi-conserved glutamate ($E^{6.53/57}$) at the ion gate (129), but cannot explain ion selectivity in K^+-dependent Na^+,H^+-PPases or K^+-independent H^+-PPases as both have the semi-conserved glutamate at $E^{6.53}$, a position consistently ascribed to Na^+-pumping (1.4.2). Based on the newly available structural data presented here, I

proposed a model for ion selectivity that focuses on the interplay of the semi-conserved glutamate with a highly conserved serine on helix 5, $S^{5.43}$ (pairwise identity: 90.6%). This updated model provides a general description of the ion dependence for all M-PPases including K^+-independent H^+-PPases and K^+-dependent Na^+,H^+-PPases (Figure 4.12).

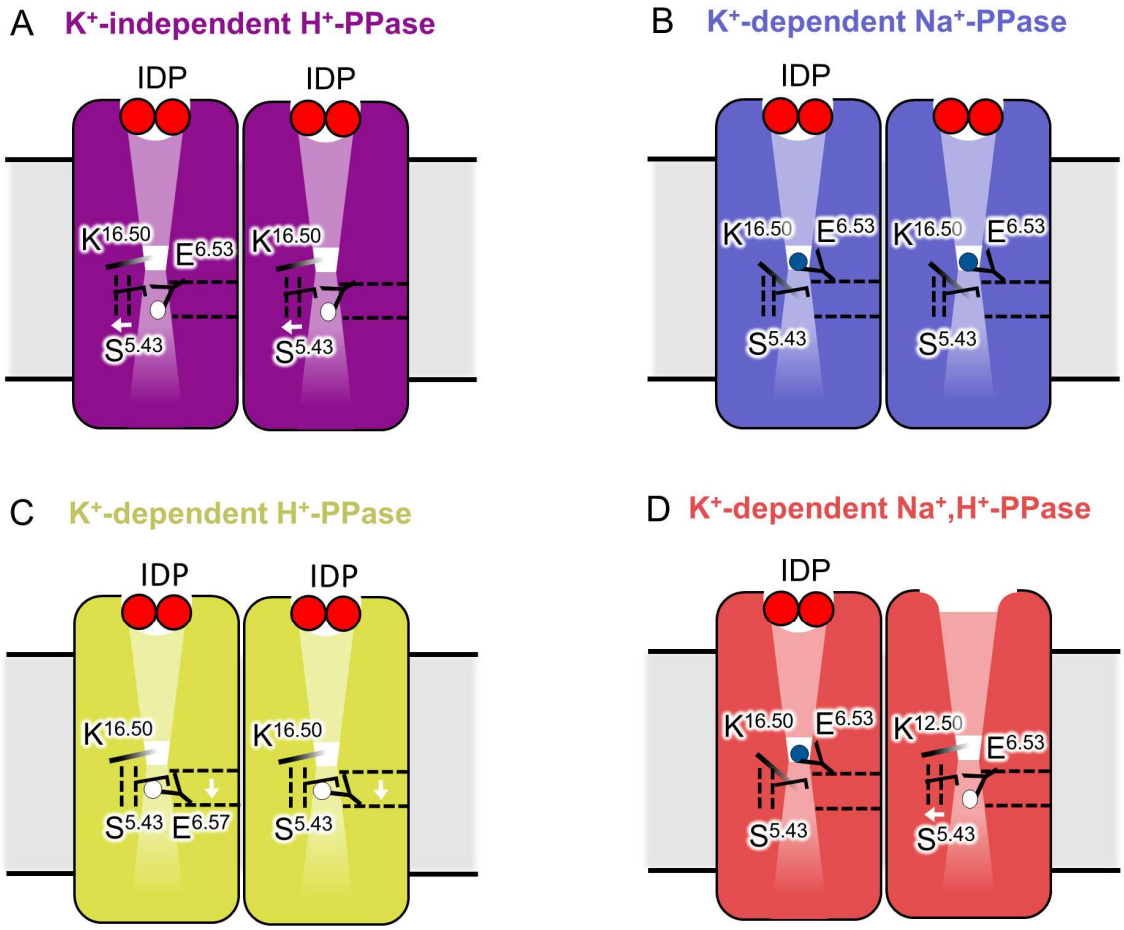


Figure 4.12: Schematic model of ion selectivity in M-PPases. The orientation of key residues $K^{16.50}$ (|) $E^{6.53/57}$ (Y) and $S^{5.43}$ (ℸ) is shown at the ion gate for each M-PPase subclass. The position of helix 5 and the semi-conserved glutamate are indicated by dashed vertical and horizontal lines, respectively. Conformational changes or residue repositioning are indicated by a white arrow. (A) $E^{6.53}$ protonation (white sphere) and interaction with $S^{5.43}$, made possible by the outward movement of helix 5. $K^{16.50}$ destroys the Na^+ binding site. (B) Na^+ (purple sphere) binding to the ion gate and interaction with $E^{6.53}$, which faces away from $S^{5.43}$ when helix 5 is close to avoid clashing. (C) $E^{6.57}$ protonation as a results of its shift one helix turn down, which creates sufficient space for the $S^{5.43}$-$E^{6.53}$ interaction. (D) Functional asymmetry at the ion gate of in K^+-dependent Na^+,H^+-PPases, explaining Na^+ binding and $E^{6.53}$ protonation in the respective subunit. Ion selectivity depends on helix 5 orientation and $E^{6.53}$-$S^{5.43}$ interaction. Subunit A is similar to the residue and helix orientation shown in panel A, whereas subunit B is similar to the residue and helix orientations of panel B.

In K^+-independent H^+-PPases such as *Pa*-PPase, $E^{6.53}$ is held in place by $S^{5.43}$, points away from the Na^+-binding site, which becomes occupied by $K^{16.50}$, and is likely protonated to reduce the negative charge in the membrane-spanning protein section (Figure 4.12A). This ion gate configuration is directly linked to the orientation of helix 5, which determines whether $E^{6.53}$ points towards $S^{5.43}$, accommodating a proton, or away from $S^{5.43}$, forming a Na^+-binding site with $D^{6.50}$, $S^{6.54}$ and $D/N^{16.46}$. For example, the bent and predominantly α-helical conformation of helix 5 in *Tm*-PPase:Mg_5IDP

requires $E^{6.53}$ to point away from $S^{5.43}$ to avoid clashing, therefore rendering the protein as a Na^+-pump (Figure 4.12B). In contrast, the straightened and tightly-wound (3_{10}-rich) conformation in *Pa*-PPase:Mg_5IDP creates sufficient space for $S^{5.43}$ to coordinate $E^{6.53}$ and prevent the formation of a Na^+-binding site, making the protein a H^+-pump (Figure 4.12A). A similar residue configuration at the ion gate might also be energetically more favourable at sub-physiological Na^+ concentrations in K^+-dependent Na^+-PPases, which would explain reported H^+-pumping activities at Na^+ concentrations <5 mM (259).

In K^+-dependent H^+-PPases, the evolutionary repositioning of the semi-conserved glutamate one helix turn down ($E^{6.57}$) uncouples ion selectivity from the orientation of helix 5 as its shift allows coordination by $S^{5.43}$, even when helix 5 does not move out to create extra space (Figure 4.12C). Similarly, the repositioning of the semi-conserved glutamate to 5.43, which is only seen in *Flavobacterium johnsoniae* M-PPase (*Fj*-PPase), renders the enzyme as an exclusive H^+-pump in line with our proposed model (260). This is because the side chain carboxylate of $E^{5.43}$ cannot promote Na^+-binding about 5 Å away, but can be modelled to occupy a position similar to that of $E^{6.57}$ in *Vr*-PPase:Mg_5IDP and accommodate a H^+ (261).

Finally, the updated model also explains ion selectivity in K^+-dependent Na^+,H^+-PPase with $E^{6.53}$ when taking functional asymmetry between subunits at the ion gate into account. In other words, in K^+-dependent Na^+,H^+-PPase the ion gate configuration of subunit A may resemble the K^+-independent H^+-PPase structure (*Pa*-PPase:Mg_5IDP), leading to H^+-pumping, whereas the configuration of subunit B resembles the K^+-dependent Na^+-PPase structure (*Tm*-PPase:Mg_5IDP), leading to Na^+-pumping, and *vice versa* (Figure 4.12D). This idea is in line with recent studies that show functional asymmetry at the active site (131,136) and there is no reason why this should not extend to the ion gate, especially due to the tight coupling between both regions. Indeed, the structural data obtained for *Pa*-PPase already reveals a plausible mechanism of intra-subunit communication between ion gates of dimeric M-PPases. The changes observed at helix 5 coincide with changes at helix 13, which is part of the dimer interface and relays structural information between subunits. Therefore, I propose that conformational changes at helix 5 in subunit A, such as straightening and tight winding with consequences for ion selectivity, are propagated into the dimer interface and subunit B *via* helix 13, where they in turn mediate helix 5 orientation and ion selectivity. The idea of functional asymmetry in M-PPases is investigated in more detail in time-resolved structural studies in Chapter 6.

Angular and linear motions like sliding, tilting or bending of helices are known to commonly influence protein function, not only in M-PPases (129). The transition between α and 3_{10} helical arrangements is less common but has been observed for other

membrane proteins such as voltage-gated ion channels (262–264). For example, the S4 helix of the ancestral bacterial sodium channel (Na_VAb) voltage sensing domain was shown to undergo a 3_{10}-α-3_{10} transitions during activation, leading to the translocation of gating charges (262). The smaller radius of a 3_{10}-helix allows more advantageous side chain orientations in S4 and fewer clashes in densely packed protein regions (262). In M-PPases, it appears that α to 3_{10} transitions around the ion gate (and straightening of helix 5) affect protein function in a similar way as they dictate whether there is sufficient space for $E^{6.53}$ to interact with $S^{5.43}$ (H^+-pumping) or not (Na^+-pumping).

4.3.6. Caveats of M-PPases activity measurements

To date, were are still missing mechanistic insights into substrate inhibition and lack an comprehensive understanding of the role of the second conserved key change, $A/G^{12.49}T$, in K^+-independent M-PPases (136). The functional studies I conducted with wild-type and variant *Pa*-PPase that aimed to better understand these processes were hampered by a combination of Mg_2PP_i solubility issues and high P_i background readings in quantitative P_i-release assays. These issues likely explain the poor fit of the obtained data to the established „half-of-the-sites“ reactivity model for M-PPases (131,136).

A recent publication from M-PPase enzymologist Dr. Aleksandr Baykov (Lomonosov Moscow State University) and co-workers highlighted some additional issues with the experimental design of hydrolytic activity assays here (265). For example, measuring the $K_{m,1}$ (~1-39 μM) and $K_{m,2}$ (~35-800 μM) of substrate binding in M-PPases requires a broader range of assayed substrate concentrations than used, particularly at the lower end (136). Moreover, continuous-flow measurements are better suited than the fixed-time assays used as substrate consumption becomes significant when working with low substrate concentrations, which violates the $[S]>>[E_{total}]$ assumption in kinetic analysis (265).

4.3.7. Energy coupling of PP_i hydrolysis and ion translocation

Unfortunately, the *Pa*-PPase structure did not provide mechanistic insights into the energy coupling mechanism of PP_i hydrolysis and ion translocation, which is disputed in the scientific community (1.5.1). This is because the obtained structure mimics reaction intermediates that can describe both mechanisms as do previously solved M-PPase structures. The discussed mechanisms mainly differ in the role of the H^+ generated during the nucleophilic attack on PP_i and, as a result of that, in the chronological order of catalytic events (1.5.1).

4.4. Future directions

Although initial efforts towards a high-resolution *Pa*-PPase structure were cut short by the SARS-CoV-2 pandemic, careful refinement and analysis an old 3.8 Å dataset yielded valuable mechanistic insights into M-PPase biochemistry. I provided the first structural evidence for $K^{12.46}$ replacing K^+ in K^+-independent M-PPases and derived an updated model of ion selectivity, which does not suffer from limited validity anymore but holds when applied to all M-PPase subclass. This significantly advances our knowledge of the molecular mechanisms defining key functions in M-PPase catalysis and is valuable information to guide the future design of experiments investigating M-PPase function or drug design to develop novel therapeutic M-PPase inhibitors.

A set of experiments will be continued by researchers at the University of Leeds and University of Helsinki to follow up on the findings of this study. These aim to further improve diffraction data quality by *in meso in situ* serial crystallography experiments at i24 (Diamond Light Source) for a more robust structure interpretation. This is complimented by atomistic MD simulations of wild-type and variant *Pa*-PPase, which were already set up by our collaborator Dr. Antreas Kalli (University of Leeds). Furthermore, functional assays testing the ion selectivity of *Pa*-PPase upon $S^{5.43}$ mutagenesis are currently being carried out by Dr. James Hillier. This includes changing $S^{5.43}$ into bulkier residues to occupy the space gained by helix 5 movement and trigger $E^{6.53}$ reorientation, which should change ion selectivity according to the updated model (4.3.5).

Finally, the experimental design of quantitative Pi-release assays needs to be adapted to investigate M-PPase kinetics more reliably. In future, the assay will be optimised to run below 40 °C and ~800µM Mg_2PP_i to prevent Mg_2PP_i precipitation (personal communication with Dr. Baykov). The detection of M-PPase hydrolytic activity at low substrate concentrations and below-optimum temperature will require the usage of malachite-green in the quantitative P_i-release assays to make it much more sensitive (265). Therefore, pyrophosphoric acid is unsuitable as source of Na^+/K^+-free PP_i in future assays due to its tendency to auto-hydrolyse in aqueous solutions and give high P_i background readings (266). Instead, cations of Na_4PP_i/K_4PP_i must be replaced on strong ion-exchange columns using weak to moderate bases, for example triethylamine (pK_a = 10.75) (personal communication with Dr. Baykov). Strong bases such as TMA-OH should be avoided as some batches were found to inhibit M-PPase activity in the past, possibly due their potential to react with glass and dissolve inhibiting ions such as Ca^{2+} (265).

Chapter 5. Towards a K^+-dependent Na^+,H^+-PPase structure

5.1. Introduction

With the availability of a K^+-independent H^+-PPase structure, dual-pumping M-PPases are the only subclass without any structural data available so far. Their structural characterisation would help to answer the remaining open questions of ion selectivity. Recently, reports of substrate inhibition upon PP_i binding to both active sites (136) and asymmetric non-phosphorus inhibitor binding to *Tm*-PPase (131) spotlighted the importance of functional asymmetry in M-PPases (1.5.2). Asymmetry in catalysis is also favoured by our updated model of ion selectivity, which links dual-pumping to asymmetric conformations of helix 5 at the ion gate and helix 13 at the dimer interface (Chapter 4). To support this idea structural data on K^+-independent Na^+,H^+-PPases is essential.

5.1.1. Summary of previous work on *Cl*-PPase

The structural characterisation of K^+-dependent Na^+,H^+-PPases has been investigated by our workgroup for many years but did not yield yet any high-resolution structures. It started in 2014 with former group member Dr. Nita Shah who investigated bacterial expression of two Na^+-regulated K^+-dependent Na^+,H^+-PPases (M-PPase from *Clostridium leptum* (*Cl*-PPase) and *Clostridium lentocellum (Cc-PPase)*) and three true K^+-dependent Na^+,H^+-PPases (M-PPases from *Akkermansia municiphila* (*Am*-PPase), *Bacteroides vulgatus* (*Bv*-PPase), *Prevoltella oralis* (*Po*-PPase)) in a range of different bacterial expression strains (BL21 DE3 pRARE2, BL21star DE3 pRARE2, C41, C43) and media (LB, M9, TB, auto-induction). In general, M-PPase expression seemed to be toxic to cells, which showed in rapid cell death upon induction. After optimisation of the expression strategy, *i.e.* inducing at very high OD_{600} at 18° C for a short time period, *Cl*-PPase, *Bv*-PPase and *Am*-PPase could be obtained in quantities sufficient for structural studies. Dr. Nita Shah also optimised the protein purification and set up the first crystallisation trials. She tested different tags (MBP, His, Strep) and on-column HRV3C cleavage protocols (protein:protease ratio, incubation time), which yielded sufficiently pure protein for crystallisation after tag removal. The choice of detergent for solubilisation and purification was mainly based on maximal protein yields and activity, with DDM, DM, Cymal6, UTM, DTM and UDM performing best out 20 detergents tested. *Cl*-PPase crystals were obtained by vapour diffusion crystallisation but only diffracted up to ~7 Å at best, even after optimisation, which was not sufficient for structure solution. LCP crystallisation was also not successful as wild-type *Cl*-PPase precipitated when

concentrated >15 mg mL^{-1}. Typically, LCP crystallisation requires higher protein concentrations (> 20 mg mL^{-1}), as the sample gets diluted with lipid (67).

5.1.2. Aims and strategy

I continued to work towards the structural characterisation of *Cl*-PPase. Previously, the selection of detergent focused on protein yield/stability and activity to obtain sufficient amounts for subsequent X-ray crystallography. Consequently, milder long-chain detergents such as DDM, UTM and DM were primarily used, but did not give good diffracting crystals. Here, I tested whether sufficiently high yields of pure protein can be obtained in harsher short-chain detergents, which are better suited for protein crystallisation by vapour-diffusion, when working with thermostabilised *Cl*-PPase variants F20Y and G130A (Chapter 3). Moreover, stabilised *Cl*-PPase might also behave better in DDM and allow its concentration above 15 mg mL^{-1} for LCP crystallisation. Ideally, these approaches will lead to high-resolution structures of *Cl*-PPase in different conformational states, *e.g.* substrate-analogue-bound (IDP), product-bound (P_i), resting (Ca^{2+}), and yield first structure-function insights into K^+-dependent Na^+,H^+-PPases.

5.2. Results

5.2.1. Heterologous protein expression and Nickel-NTA purification of wild-type and thermostabilised *Cl*-PPase

The purification of wild-type and variant (F20Y, G130A) *Cl*-PPase followed a protocol developed by Dr. Nita Shah (2.5.1). *Cl*-PPase was expressed in BL21 DE3 star pRARE2 cells during a short 1-hour IPTG induction at low temperature (18 °C) and high cell densities (OD_{600}: 6-7). Different types of TB media affected the growth curves of expression strains significantly. The obtained bacterial cell mass was ~2-fold in self-made media compared to commercially available media from Alfa Aesar (Figure 5.1A), for which no recipe is published. The expression levels of *Cl*-PPase variants F20Y and G130A were within ~7% of wild-type *Cl*-PPase (Figure 5.1B). Similarly, the efficiency of solubilisation in DDM (87-90%) and cobalt resin-binding (84-97%) efficiency were also almost identical between wild-type and variant protein (Figure 5.1C). Tag removal by HRV3C cleavage and subsequent elution from the purification column was less successful. Most of the target protein remained uncleaved after 4 hours digestion at 4 °C (1:1000 protease to protein) or did not elute from the resin (Figure 5.1C), leaving room for further optimisation. Nevertheless, the yield of purified protein was ~1.5-fold (0.5-0.75 mg L^{-1}) higher than what was obtained in previous studies (5.1.1) and sufficient for crystallisation. I did not follow up on the expression of *Cl*-PPase in *S. cerevisiae* as yields did not exceed 0.1 mg L^{-1} culture in FGY217 (used in *Cl*-PPase stability studies)

and Dr. Nita Shah previously reported poor expression in yeast strain BJ1991 (used for *Pa*-PPase and *Tm*-PPase production).

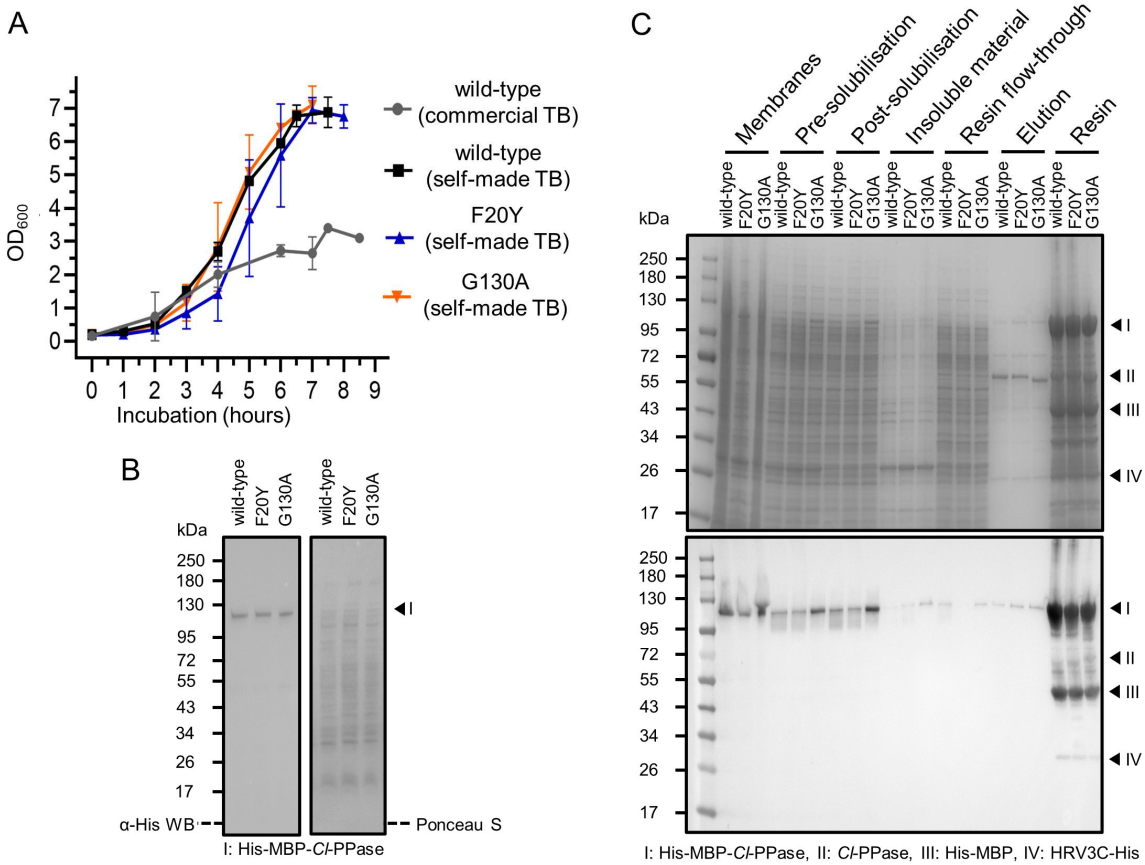


Figure 5.1: Wild-type and variant *Cl*-PPase expression and purification. (A) BL21star growth curves in TB media. *Cl*-PPase expression was induced when cells approached the stationary phase (after 6-7 hours incubation) with 0.5 mM IPTG. (B) Expression level comparison of wild-type and variant *Cl*-PPase by densiometric analysis of His-MBP-*Cl*-PPase bands from SDS-treated cell sample. Left panel shows α-His-WB. Right panel shows Ponceau S staining of PVDF membrane for protein loading correction. (C) Small-scale purification of wild-type and variant *Cl*-PPase alongside each other. Upper panel shows SDS-PAGE analysis (Coomassie) of samples taken throughout the purification. Lower panel shows the corresponding α-His-WB. Bands of interest are indicated by arrows and labelled below the panels.

SDS-PAGE analysis of *Cl*-PPase purification products showed that the G130A variant migrated slightly below the F20Y variant and wild-type *Cl*-PPase (Figure 5.1C). Proteolytic degradation was ruled out by LC-MS (Mass Spectrometry Facility at the University of Leeds), confirming the expected mass for all three *Cl*-PPase variants tested (Figure 5.2). The mass difference between wild-type, F20Y and G130A variants fits the expected shift caused by the amino acid substitutions (Table 5.1). In addition to unmodified full-length *Cl*-PPase, each sample was contaminated with a less abundant modified *Cl*-PPase species (mass addition of 320-324 Da), which co-eluted with all samples in LC. The cause of this mass difference remains unknown as non-covalently linked molecules are stripped off during LC-MS and no post-translational modifications fit the observed mass addition, nor are any known for bacterial M-PPases.

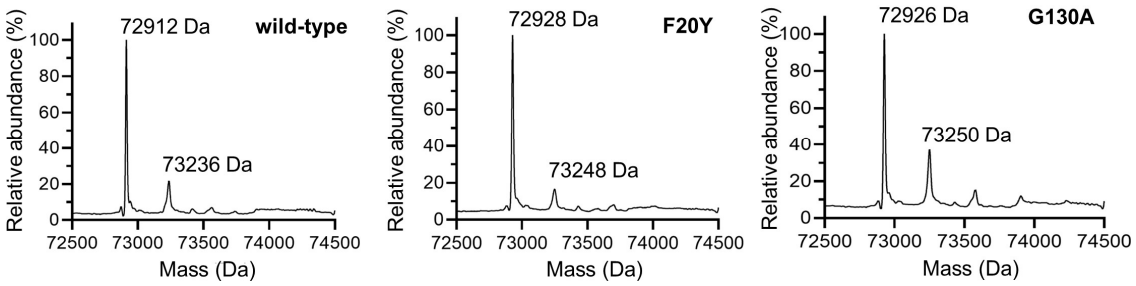


Figure 5.2: Mass measurements of wild-type and variant *Cl*-PPase by LC-MS. Deconvoluted mass spectrum of protein peak eluting at ~7.5 minutes in LC.

A few contaminating bands were present in the IMAC elution fraction, mainly uncleaved His-MBP-*Cl*-PPase, cleavage products and HRV3C (Figure 5.1C). This improved in large scale purification batches (10 L) of DDM-solubilised protein, which provided sufficient quantities (~5 mg) for subsequent crystallisation trials (Figure 5.3A). The sample purity was comparable to *Pa*-PPase and *Tm*-PPase purification batches, which both readily crystallise after IMAC elution (Chapter 4, Chapter 6). Previously, *Cl*-PPase also crystallised without additional purification (5.1.1). Therefore, *Cl*-PPase variants were used for LCP crystallisation straight after on-column tag-removal, protein elution and concentration (Figure 5.3A).

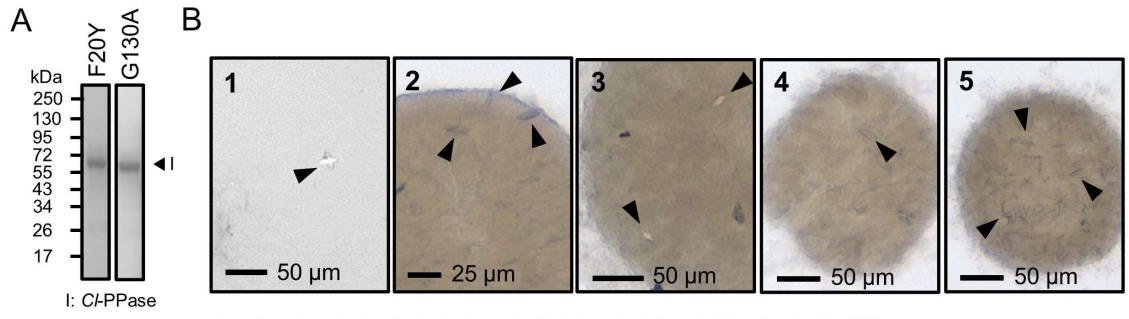
Table 5.1: Table of expected and measured masses from the most abundant species in LC-MS.

Mass (Da)	*Cl*-PPase variants		
	wild-type	F20Y	G130A
expected	72912	72928	72926
measured	72912	72928	72926
Δ mass	± 0	± 0	± 0
measured	73236	73248	73250
Δ mass	+ 324	+ 320	+ 324

5.2.2. LCP crystallisation trials of *Cl*-PPase variants F20Y and G130A

LCP crystallisation typically requires protein concentrations of >20 mg mL^{-1}, the higher the better for initial screening (67). Wild-type *Cl*-PPase could not be concentrated to >15 mg mL^{-1} and protein precipitated on ice over time. Although the concentration limit for *Cl*-PPase F20Y was 15-20 mg mL^{-1}, the protein was stable on ice for several hours and could be co-crystallised with the non-hydrolysable inhibitor IDP and reconstituted into LCP. *Cl*-PPase G130A, which is more stable than the F20Y variant (Chapter 3), could be concentrated to 28 mg mL^{-1}. Crystal screening focused on conditions that yielded *Pa*-PPase crystals in LCP with 9.7MAG as host lipid and in the presence of IDP (4.2.2). Moreover, several commercially available sparse-matrix screens were set up for both variants. The most promising conditions identified in screening are shown in Figure 5.3B. These crystals were harvested and tested at microfocus beamlines. However, no diffraction was observed in mesh scans. It cannot be ruled out that crystals were lost or destroyed during the harvesting process as LCP turned opaque upon opening of the well. Alternatively, these crystals might have been salt or detergent instead of protein. Unfortunately, I could not pre-examine crystals by UV-light absorption

tests to confirm the presence of protein prior to diffraction experiments due to a faulty camera system.



1. 0.1 M MES-NaOH pH 6.5, 0.2 M $MgCl_2$, 25 % *v/v* PEG 400
2. 0.1 M MES-NaOH pH 6.0, 0.1 M NaCl, 0.1 M $LiSO_4$, 40 % *v/v* PEG 200
3. 0.1 M HEPES-HCl pH 7.0, 30% *v/v* PEG 500 DME
4. 0.1 M sodium citrate pH 5.0, 0.1 M NaCl, 0.1 M L_iSO_4, 30% *v/v* PEG 300
5. 0.1 M Tris-HCl pH 8.0, 0.1 M $CaCl_2$, 30% *v/v* PEG 400

Figure 5.3: LCP crystallisation of *Cl*-PPase. (A) Coomassie-stained SDS-PAGE analysis of elution fraction from large scale purification of F20Y and G130A in DDM. Protein used in subsequent LCP crystallisation trials. (B) Promising hits of LCP crystallisation of *Cl*-PPase$_{G130A}$ in complex with IDP and 9.7MAG as host lipid.

5.2.3. Detergent screening

A second approach to improve diffraction quality of crystals focused on the use of short-chain detergents (<10 carbon atoms in alkyl chain) in vapour-diffusion crystallisation trials, which may interfere less with crystal contact formation. Therefore, I quantified the purification yields in a range of short-chain detergents using wild-type and thermostabilised *Cl*-PPase variants and compared these to protein purified in DDM (0.5-0.75 mg L^{-1} culture) (Figure S7, Appendix). As expected, the stabilised *Cl*-PPase variants F20Y and G130A are more “detergent-resistant” than wild-type *Cl*-PPase and could thus be purified in harsher detergents such as OGNG, NG, OM or OG (Figure 5.4A). Moreover, *Cl*-PPase clearly favoured detergents with an alkyl chain length ≥9 (2-fold yields in NM compared to OM) and β-D-maltoside head groups instead of β-D-glucoside head groups (3-fold yields in NM compared to NG). Therefore, NM was chosen for future crystallisation trials. Larger expression batches may even yield enough protein to test crystallisation in OM as well.

Prior to crystallisation trials, *Cl*-PPase activity in DDM and NM was assessed using the quantitative P_i-release assay. First, the incubation time was adjusted to ensure a linear P_i-release with time at 30 °C using wild-type protein solubilised in DDM (Figure 5.4B). After ~15 minutes half of the available substrate in the reaction condition was turned over, which translates to a specific activity of 3.8 ±0.3 µmol P_i mg^{-1} min^{-1}. In NM, the activity was reduced 4-fold but protein remained active (Figure 5.4C). The effect of the thermostabilising mutations F20Y and G130A on hydrolytic activity was only investigated for *Cl*-PPase purified in DDM and found to be similar to earlier activity measurements of wild-type and variant *Cl*-PPase in the membrane (3.2.1.2). The G130A change affected hydrolytic activity less (67% compared to wild-type) than F20Y (165% compared to

wild-type), which even improved substrate turnover (Figure 5.4C). Therefore, it is likely that *Cl*-PPase variants will also be active in NM. This makes NM a prime candidate for future vapour-diffusion crystallisation trials (5.4).

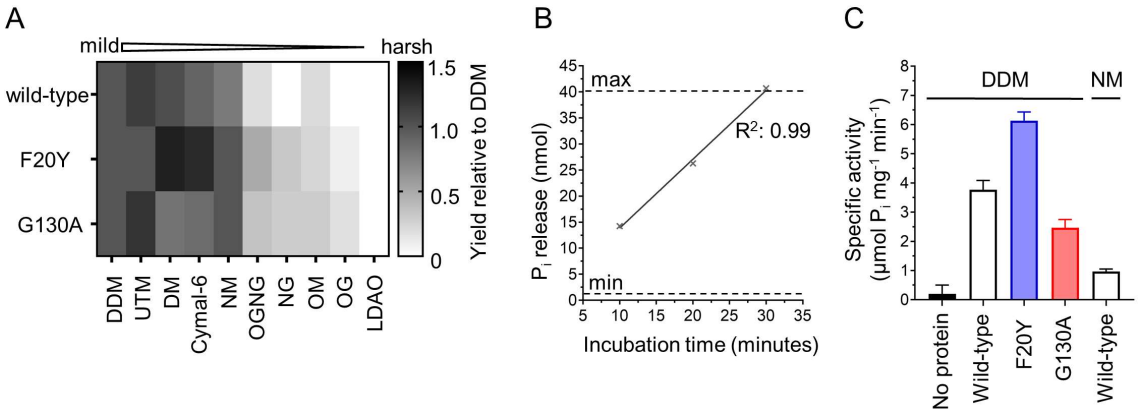


Figure 5.4: Small-scale detergent screening. (A) Purification yields of wild-type and variant *Cl*-PPase in increasingly (left to right) harsh detergents in relation to protein amount obtained in DDM. (B) Quantitative P_i-release activity assay optimisation to ensure P_i release by wild-type *Cl*-PPase is linear with time during incubation at 30 °C. Dashed lines indicate minimal (min) P_i release detectable and maximal (max) P_i release if all substrate is turned over. Linear regression and coefficient of determination (R^2) shown. (C) Specific activity (µmol mg^{-1} min^{-1}) of wild-type and variant protein purified in different detergents (DDM or NM).

5.3. Discussion

In this study, I investigated the potential of the two most stable *Cl*-PPase variants identified in a rationalised screening campaign in Chapter 3, F20Y and G130A, to yield good-quality protein crystals. Their behaviour was assessed throughout the whole protein production process to the point of crystallisation in LCP or *via* the vapour-diffusion method. Both mutations likely stabilise *Cl*-PPase by rigidifying dynamic protein regions, particularly helix 1 (3.3.1).

Wild-type and variant protein were almost indistinguishable during protein expression, membrane extraction and IMAC purification in mild detergent. The 1.5-fold increase in purification yields of wild-type and variant *Cl*-PPase is likely the result of the increased cell mass produced in self-made TB media as the solubilisation and tag removal efficiency were comparable to Dr. Nita Shah's results (not published). In commercial TB media, cell growth stagnated at much lower cell density. The commercial recipe is not published but comparison to the most used TB recipes revealed a difference in pH by 0.5 units (commercial: pH 7.5, self-made: pH 8.0). It remains unclear why this improved cell growth in shaking cultures or whether the commercial receipt differed in other ways. Throughout the purification process, thermostabilised *Cl*-PPase variants only gave a

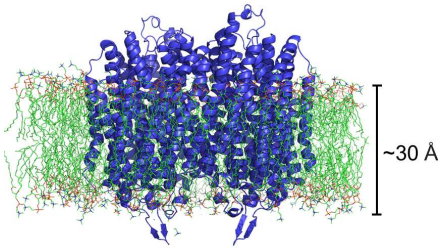


Figure 5.5: *Tm*-PPase in a DPPC membrane. Final frame of a 1 µs Coarse-grained MD simulation after conversion to atomistic resolution. The membrane spanning protein section measures ~30 Å.

significant advantage over wild-type protein when working under less optimal conditions, for example at the high protein concentrations required for LCP crystallisation (62) or in harsher short-chain detergents that are more favourable for vapour-diffusion crystallisation (267). Interestingly, there was a clear cut-off for the detergent alkyl chain length below which purification yields decreased significantly. Coarse-grained MD simulations of *Tm*-PPase and *Vr*-PPase in a lipid bilayer (268) show that the membrane spanning protein section covers ~30 Å in M-PPases (Figure 5.5). This correlates with the distance between opposing head groups in detergent-micelles with an alkyl chain length of 9 carbon atoms, such as NM (13). In shorter-chain detergents ($<C_9$) hydrophobic protein patches are likely exposed leading to increased aggregation and protein precipitation. For example, OM micelles only bridge a distance of ~28 Å between opposing head groups (13). The rationale of increased protein stability in β-D-maltoside detergents compared to β-D-glucoside detergents is less obvious and remains unclear in the absence of high-resolution *Cl*-PPase structures that resolve interaction with detergent molecules.

Overall, the introduction of thermostabilising point mutations improved protein handling and purification in harsh detergents, which enables *Cl*-PPase crystallisation in LCP or short-chain detergents (*e.g.* NM). Due to the SARS-CoV-2 pandemic limiting access to research facilities, I could not set up vapour-diffusion crystallisation trials with protein purified in NM nor optimise crystals obtained in LCP. Although LCP crystals were not confirmed to contain protein, it is more likely that the absence of X-ray diffraction was caused by issues during harvesting (5.2.2) as the tested crystals were obtained based on conditions that yielded diffracting *Pa*-PPase crystals in LCP previously (4.2.2). The issues of crystal harvesting from LCP might be overcome by growing more and larger crystals or by following an *in meso in situ* approach, which allows crystal diffraction from unopened wells (256).

5.4. Future directions

Future experiments will focus on more extensive crystallisation trials with thermostabilised *Cl*-PPase variants F20Y and G130A. LCP crystals should be reproduced in IMISX™ plates to confirm protein diffraction and subsequently optimised in finer salt and PEG screening. Additionally, co-crystallisation with a broader range of M-PPase inhibitors including non-phosphorous agents that were recently identified in drug discovery screening should be tested (131,168). Complementary vapour-diffusion crystallisation approaches may focus on protein purified in NM. Harsher detergents such as OM might require further stabilisation of *Cl*-PPase by combining stabilising point-mutations. As pointed out earlier, this is not trivial and might require systematic and laborious screening (3.3.1). Fractional factorial design can help to reduce the

experimental burden of this by carefully selecting a subset of all possible combinations, which would correspond to a full factorial design, to exploit the sparsity-of-effects principal and identify key determinants of protein stability (269). Purification in harsher detergents might also require increased protein yields to cover for protein loss due to instability. Currently, most of the protein is not cleaved off the column by HRV3C digestion. Therefore, the protease to protein ration and incubation time should be further optimised. These directions are currently being explored by Dr. James Hillier.

Chapter 6. Time-resolved structural studies of K^+-dependent Na^+-PPase

6.1. Introduction

More and more evidence indicates functional asymmetry in M-PPases, for example: kinetic data showing substrate inhibition upon occupation of both active sites (136), asymmetrical loop orientations trapped by a non-phosphorous inhibitor (131) and the structural characterisation of an K^+-independent H^+-PPase giving rise to an updated model of ion-selectivity, which can explain dual-pumping in M-PPases by asymmetric helix 5 orientations (4.3.5). However, mapping of physiologically relevant asymmetric enzyme conformations proved very difficult using common crystallographic approaches. Competitive inhibitors that are routinely used for conformational stabilisation typically occupy both active sites and are thought to remove any functional asymmetry previously present in M-PPases (129). Here, I worked towards mapping more native like asymmetric enzyme conformations by time-resolved X-ray crystallography without using any inhibitors.

6.1.1. Concept of time-resolved X-ray crystallography

In contrast to conventional X-ray crystallography that provides static snapshots, time-resolved X-ray crystallography probes active protein during its reaction cycle after a time-delay to capture structural dynamics for a more complete and in-depth understanding of the biological process investigated. This requires the crystallised protein to be active, which excludes the use of inhibitors to aid crystallisation by conformational stabilisation/trapping (270). Moreover, the *in crystallo* reaction initiation needs to be precisely controlled to probe protein in the same state during the diffraction experiment (270). This dictates additional requirements for time-resolved crystallography. The enzymatic turnover or average lifetime of targeted intermediate conformations needs to be slower than the reaction initiation, *i.e.* the time it takes for the reaction trigger to arrive at its site of action (270). In the case of chemical triggers, this is mainly dependent on the diffusion coefficient of activating molecules (*e.g.* substrate) and the crystal size. The latter is often adjusted for a better time-resolution (up to microseconds) if the reaction is initiated by rapid mixing (271). A much faster, almost instant (femtosecond time-scale), reaction initiation can be achieved by using laser light pulses to trigger photo-active proteins (272), induce an activating temperature jump (273) or release photo-caged compounds (274). Probing or stopping the reaction after a certain time-delay is equally important. This can be done by crystal flash cooling in liquid nitrogen, which is the easiest method and allows the collection of a complete dataset

from a single crystal. However, manual crystal handling takes seconds, which is too slow to capture most protein motions (275). Much better time-resolutions can be obtained following a serial crystallography approach in which hundreds to thousands of crystals are shot once after a certain time-delay to capture enough orientations for a complete dataset (275). Serial time-resolved crystallography experiments can capture biological processes with a time-resolution down to picoseconds at specialised synchrotron beamlines (276) or femtoseconds at free electron laser (X-FEL) beamlines (277). The crystals are typically delivered into the X-ray beam on fixed-targets (*e.g.* chips) (192), by high-viscosity injectors (278) or in micro-fluidic devices (279), all having distinct advantages and disadvantages discussed elsewhere (270,275,277). On chips, the time-delay between reaction initiation and probing can be controlled in a hit-and-return approach (HARE) by activating (rapid mixing or photoactivation) a set of crystals and returning to these for diffraction data collection after a defined period of time (192). In micro-fluidic devices (rapid mixing) and high-viscosity injectors (photoactivation), the time-delay is controlled by the time of travel between the point of reaction initiation and the X-ray beam and can be adjusted by changing the distance between the two (278,279). Due to the many additional requirements, for example reaction initiation and stopping, time-resolved X-ray crystallography is very challenging. Nevertheless, the sheer amount of information and in-depth insight into the molecular mechanism of biological processes that time-resolved structural studies offer, and the comprehensive support provided at specialised beamlines encourages a growing number of scientists to enter the field.

6.1.2. Aims and Strategy

M-PPase are known to be very slow enzymes. Their reaction kinetics are up two to orders of magnitude slower (catalytic substrate turnover: 10-100 s^{-1}) than soluble pyrophosphatases (73). This is sufficiently slow to enable rection initiation by rapid mixing, which I explored by using single-crystal and serial crystallographic approaches. Serial time-resolved X-ray crystallography requires large volumes (~200 µL) of high-density crystal suspensions to collect enough crystal orientations for complete datasets at different timepoints (192). Here, I used a fixed target crystallisation chip that can hold over 20.000 individual crystals (192). To produce sufficient numbers of crystals, I crystallised in batch (Eppendorf tubes) after optimising crystallisation conditions for time-resolved studies in vapour diffusion experiments, which requires much less sample. *Tm*-PPase, a K^+-dependent Na^+-PPase, was used as test protein as it is naturally thermostable, can be obtained in milligram quantities and readily crystallises after purification. For conformational stabilisation without the use of inhibitors, I removed Na^+ from the protein buffer and crystallisation condition as ion-pumping and hydrolysis are tightly coupled in M-PPases (1.5.1). This should enable the crystallisation in presence of

physiological substrate (PP_i) and subsequent reaction initiation by adding back Na^+. I aimed to map intermediate and possibly asymmetric enzyme conformations to answer mechanic questions regarding the relay of the active site and ion gate status between subunits and the role of cationic centre in substrate inhibition.

6.2. Results

6.2.1. Heterologous protein expression and hot-solve purification

The expression and purification of *Tm*-PPase followed the "hot-solve" protocol that is routinely used for thermostable M-PPases and utilises elevated temperatures to enhance detergent-extraction and heat precipitate mesophilic host proteins, thereby improving protein yield and purity (8). About 0.5 mg L^{-1} culture of wild-type *Tm*-PPase were obtained following this method. Protein was purified in OGNG, known to give well-diffracting crystals (up to 2.6 Å) (10), from a total culture volume of 100 L (10 L batches). This yielded ~45 mg of purified *Tm*-PPase for batch crystallisation trials.

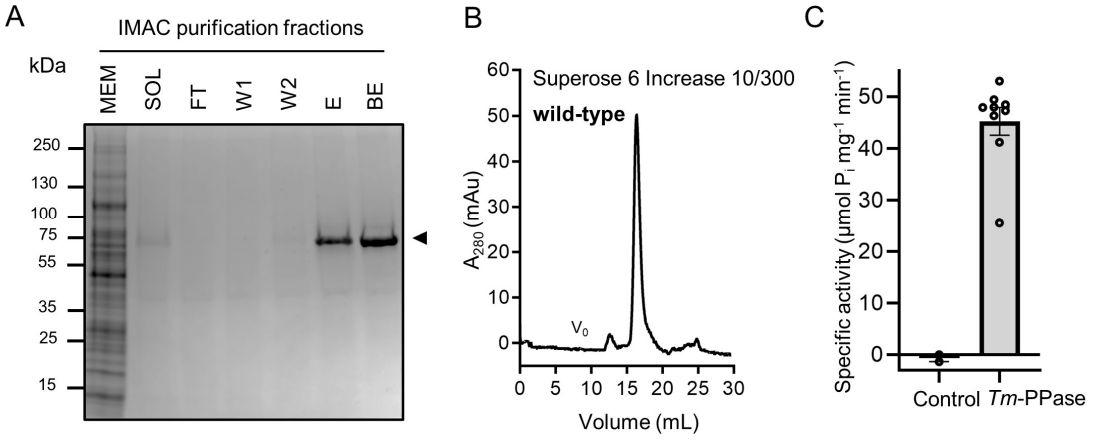


Figure 6.1: Purification and functional characterisation of wild-type *Tm*-PPase. (A) SDS-PAGE (Coomassie stain) analysis of "hot-solve" purification in OGNG. Resuspended membrane sample (MEM), detergent-solubilised crude membrane extract (SOL), column wash 1-2 (W1, W2), protein elution (E) and buffer exchanged sample (BE) loaded. Purified full-length *Tm*-PPase bands are highlighted by a black arrow. (B) Analytical SEC of wild-type *Tm*-PPase on Superose 6 Increase 10/300 column. The void volume is indicated by v_o.(C) Quantitative P_i-release activity assay to assess the hydrolytic activity of individual purification batches (each represented by a black circle). The negative control sample lacks *Tm*-PPase. Error bars denote SEM.

The inspection of the IMAC elution fractions on SDS-PAGE showed that protein was pure (Figure 6.1A), which was further confirmed by analytical SEC. The eluted protein peak was monodisperse and only showed a small shoulder towards larger elution volumes (Figure 6.1B). Eluted protein was then buffer exchanged to remove imidazole and Na^+ from the sample and concentrated to ~10 mg mL^{-1} for crystallisation. As quality control, the hydrolytic activity of each purification batch was determined using the routine quantitative colorimetric P_i-release assay that was previously optimised for thermophilic *Tm*-PPase (114). All *Tm*-PPase samples purified in OGNG were active with an average specific activity value of 45.2 ±2.7 µmol P_i mg^{-1} min^{-1} (Figure 6.1C).

6.2.2. Vapour diffusion and batch crystallisation for time-resolved structural studies

Initial crystallisation trials for time-resolved structural studies were continuously optimised to yield Na^+-free *Tm*-PPase crystals in the presence of PP_i (6.1.2) and based on a crystallisation condition with 50 mM Tris-HCl pH 8.5, 36% *v/v* PEG400, 100 mM $MgCl_2$, 100 mM NaCl and 2 mM DTT that yielded the *Tm*-PPase:Mg_5IDP structure (Figure 6.2A). This condition was chosen since its individual components are all present in the routine reaction mixture of M-PPase activity assays, except for PEG400, and their effect on protein activity is well documented (136,134). In iterative rounds of crystal optimisation, vapour diffusion conditions were adjusted to be compatible with a time-resolved approach by conformational stabilisation/locking of *Tm*-PPase in the absence of Na^+. At first, NaCl and Na_4PP_i were replaced with KCl and K_4PP_i to remove Na^+ in solution (Figure 6.2B). Afterwards, the precipitant concentration was lowered to 32% *v/v* PEG400 to give fewer but bigger and less branched crystals (Figure 6.2C). Then, the $MgCl_2$ and K_4PP_i concentration was lowered to better resemble reaction conditions of the routine hydrolytic activity assays (Figure 6.2D). As KCl is the only salt present that does not affect protein activity when kept at ≥50 mM, its concentration was gradually increased to 175 mM (Figure 6.2E) to improve crystal shape (*e.g.:* sharp edges, no branching, no stacking). Finally, the precipitant concentration was further lowered to 24-26% *v/v* PEG400 as this yielded squared-shape plates of up to 200 x 200 µm without any visible defects (Figure 6.2F). The optimisation process and individual conditions used are summarised in Table 6.1.

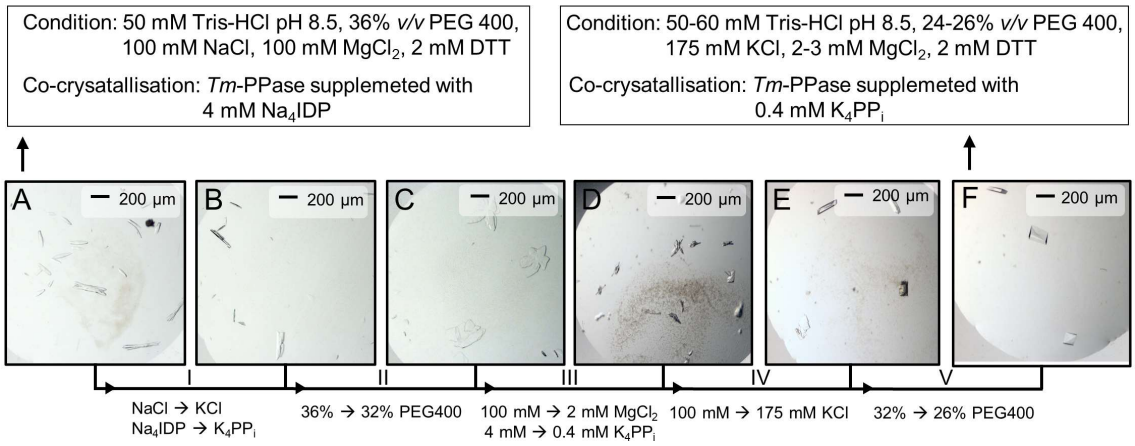


Figure 6.2: Vapour diffusion crystallisation and crystal optimisation rounds (I-V). The different conditions used for crystallisation in panels A-F are listed in Table 6.1. The start (A) and end condition (F) of optimisation are listed in black boxes

Table 6.1: *Tm*-PPase crystallisation condition optimisation for time-resolved structural studies

Optimisation	Crystallisation condition	Co-crystallisation
Original: Figure 6.2A	50 mM Tris-HCl pH 8.5, 36% *v/v* PEG400, 100 mM $MgCl_2$, 100 mM NaCl, 2 mM DTT	4 mM Na_4PP_i
Condition I: Figure 6.2B	50 mM Tris-HCl pH 8.5, 36% *v/v* PEG400, 100 mM $MgCl_2$, 100 mM KCl, 2 mM DTT	4 mM K_4PP_i
Condition II: Figure 6.2C	50 mM Tris-HCl pH 8.5, 32% *v/v* PEG400, 100 mM $MgCl_2$, 100 mM KCl, 2 mM DTT	4 mM K_4PP_i
Condition III: Figure 6.2D	50 mM Tris-HCl pH 8.5, 32% *v/v* PEG400, 2 mM $MgCl_2$, 100 mM KCl, 2 mM DTT	0.4 mM K_4PP_i
Condition IV: Figure 6.2E	50 mM Tris-HCl pH 8.5, 32% *v/v* PEG400, 2 mM $MgCl_2$, 175 mM KCl, 2 mM DTT	0.4 mM K_4PP_i
Condition V: Figure 6.2	50 mM Tris-HCl pH 8.5, 26% *v/v* PEG400, 2 mM $MgCl_2$, 175 mM KCl, 2 mM DTT	0.4 mM K_4PP_i

Two vapour-diffusion crystallisation conditions for time-resolved studies (Table 6.1, I+V) were adapted for large-scale batch crystallisation (~45 µL) to obtain high-density suspensions of crystals that can be loaded on hit-and-return (HARE) crystal support chips (192). This allows utilisation of the liquid application method for time-resolved analysis (LAMA) in which picolitre droplets of trigger solution are shot at individual crystals on the HARE chip for *in situ* rapid mixing and data collection within milliseconds (280). The transition from vapour-diffusion crystals to batch crystals followed an adapted protocol by Stohrer and co-workers (179). The concentration of precipitant, here PEG400, was increased to match the equilibrium concentration of precipitant in vapour-diffusion experiments when protein is mixed with crystallisation solution at a 1:2 ratio. This yielded batch crystals based on vapour-diffusion crystallisation condition I (Table 6.1), which grew in in 33 mM Tris-HCl pH8.5, 37% *v/v* PEG400, 67 mM $MgCl_2$, 67 mM KCl and 0.4 mM K_4PP_i. Lowering the precipitant concentration to 24% *v/v* PEG400 resulted in less branching and sharper edges (Figure 6.3A). In a second batch crystallisation trial the concentration of buffer and salts were increased alongside the precipitant to closely resemble the optimised vapour-diffusion crystallisation condition V upon mixing with protein (Table 6.1). The crystal shape was fine-tuned by lowering the precipitant concentration to 24% *v/v* PEG400. This yielded square-shaped plates that resembled the final optimised crystals of vapour-diffusion experiments (Figure 6.3B).

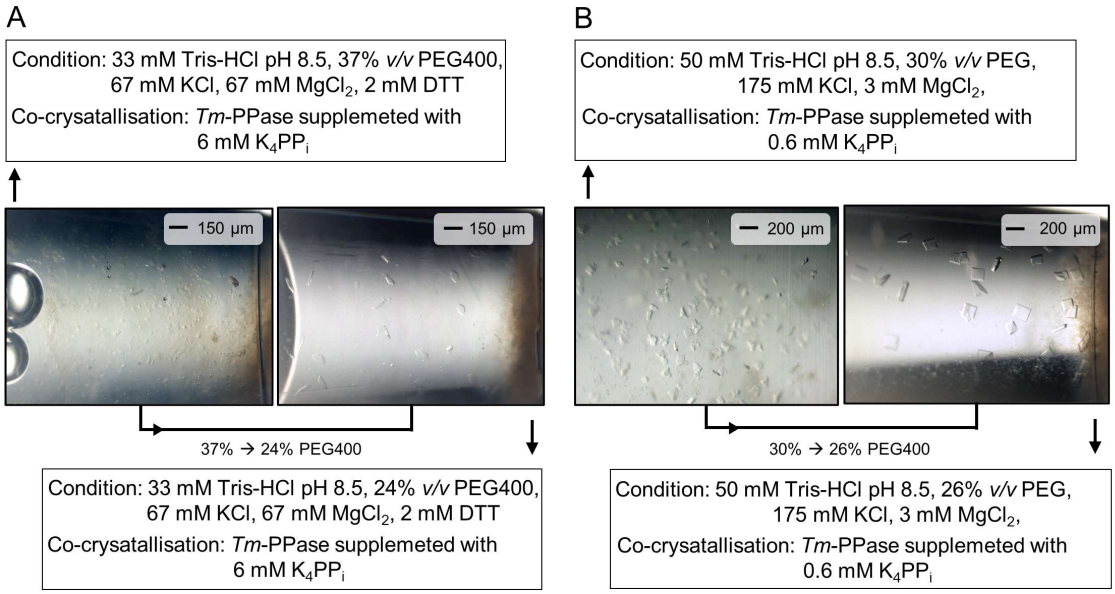


Figure 6.3: *Tm*-PPase batch crystallisation in two different conditions and crystal optimisation. (A) Batch crystallisation based on vapour diffusion condition I (Table 6.1). (B) Batch crystallisation based on vapour diffusion condition V (Table 6.1). The start and end condition of optimisation are listed in black boxes.

The diffraction power of vapour diffusion grown crystals was assessed at beamline P14-I (DESY) in single-crystal X-ray diffraction experiments at 100 K (Figure 6.4A). The X-ray diffraction of optimised crystals was highly anisotropic with reflections visible up to 2.8 Å in the best direction. The unit cell dimensions of data processed in space group $P2_1$ were 84.0 Å, 110.2 Å, 108.0 Å (a, b, c) and 107.9°(β). Crystal plates obtained in batch experiments from pre- and post-optimisation crystallisation trials were loaded into a HARE chip (one condition in each corner) and shot at 293K at beamline P14-II (DESY) (Figure 6.4B), which is dedicated to time-resolved serial crystallography experiments. The quality of data collected at P14-II was much worse as X-ray diffraction was limited to ~7 Å and too few crystals were hit to give a complete serial dataset. Moreover, data integration indicated the presence of two different crystal forms in all batch conditions with either cubic or orthorhombic lattices and unit cell dimensions of 156.1 Å or 105.1 Å, 115.3 Å and 157.5 Å, respectively.

In summary, only vapour-diffusion crystals proved suitable for time-resolved structural studies although batch crystals were obtained from similar final crystallisation conditions. This eliminated the LAMA approach on HARE chips for time-resolved structural studies as sitting or hanging drop vapour diffusion crystallisation cannot satisfy its requirement for large volumes of high-density crystal suspensions. Using single vapour-diffusion grown crystals for time resolved structural studies is only compatible with manual mixing and flash cooling in liquid nitrogen. This requires the enzymatic turnover being slow comparted to ligand diffusion, here Na^+, and manual crystal handling from the point of reaction initiation to the reaction stop. The diffusion time of glucose into vapour-diffusion crystal plates with volume of 80 mm^3 (200x200x2 µm) is ~100 milliseconds based on

Fick's second law describing the change of concentration with time (140). Although the diffusion of Na^+ is likely faster, manual crystal handling takes seconds so only intermediates with characteristic lifetimes longer than that can be resolved in time-resolved structural studies following a manual approach.

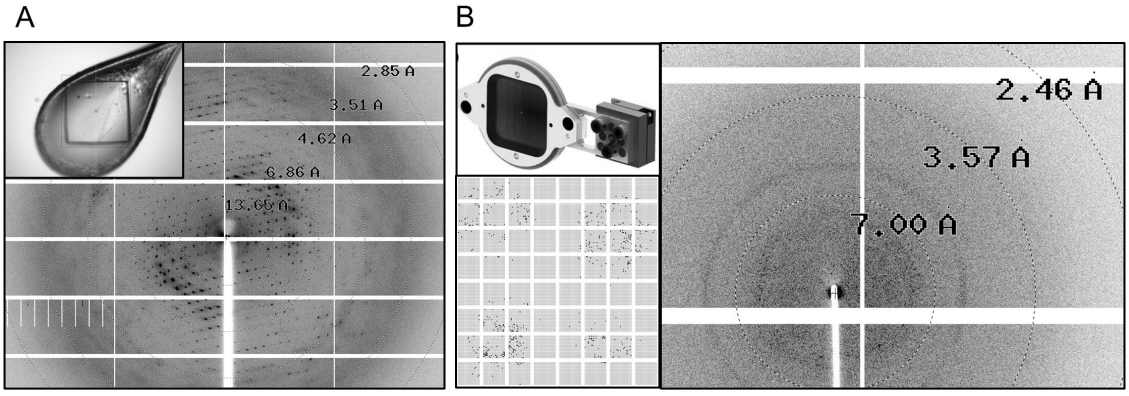


Figure 6.4: X-ray diffraction of *Tm*-PPase crystals used for time-resolved experiments. (A) Diffraction pattern of optimised vapour-diffusion crystals shot at 100 K at P14-I. Upper left panel shows the harvested crystal in a nylon loop at the beamline. (B) Diffraction pattern of batch-grown crystals shot at 293 K at P14-II. 4x45 µL crystal batches were loaded onto the corners of a HARE chip (upper left panel, adapted from Mehrabi *et al.* 2020). The lower left panel shows a representation of the crystal wells with detected diffraction from wells indicated by black dots.

6.2.3. Activity measurements in crystallisation conditions

I assessed the enzymatic turnover of PP_i by *Tm*-PPase in "crystallisation-like" conditions (Table 6.2) to evaluate the feasibility of manual rapid mixing and reaction stopping in time-resolved crystallographic studies. Therefore, the activity assay was optimised for use at 20 °C at which crystals were obtained and stable. To detect hydrolytic activity in quantitative P_i-release assays at below-optimum temperatures (20 °C instead of 71 °C), the minimum reaction time needs to be increased to ~20 minutes to compensate for the reduced substrate turnover rates (Figure 6.5A). P_i release was linear with time to up to a reaction period of 240 minutes. As *Tm*-PPase is expected to be less active in "crystallisation-like" conditions than in standard conditions (Table 6.2), all subsequent assays at 20 °C were run for 240 minutes to ensure that P_i release is sufficient for detection at non-optimal conditions.

Table 6.2: Activity assay conditions.

Reaction mixture	Condition
Standard	60 mM Tris-HCl, 100 mM KCl, 10-20 mM NaCl, 3-5 mM $MgCl_2$, 0.4 mM
Crystallisation-like	60 mM Tris-HCl, 175 mM KCl, 20 mM NaCl, 3 mM $MgCl_2$,

Initially, I confirmed that protein can be activated by Na^+ instead of PP_i in accordance with the strategy used for crystallisation. This was done using standard reaction conditions at the *Tm*-PPase temperature optimum of 71 °C. Usually, trigger solution contains K_4PP_i and is added to the reaction solution, which has all other components required for M-PPase activity including Na^+, for reaction initiation. Here, I inter-exchanged Na^+ and PP_i in the reaction and trigger solution. This showed that Na^+

activation by adding a NaCl containing trigger solution to a K_4PP_i containing reaction condition is equally suitable for *Tm*-PPase reaction initiation (Figure 6.5B). Importantly, without addition of NaCl no activity was detected throughout all assays so trace amounts of Na^+ were not sufficient to trigger significant substrate turnover (Figure 6.5B-E).

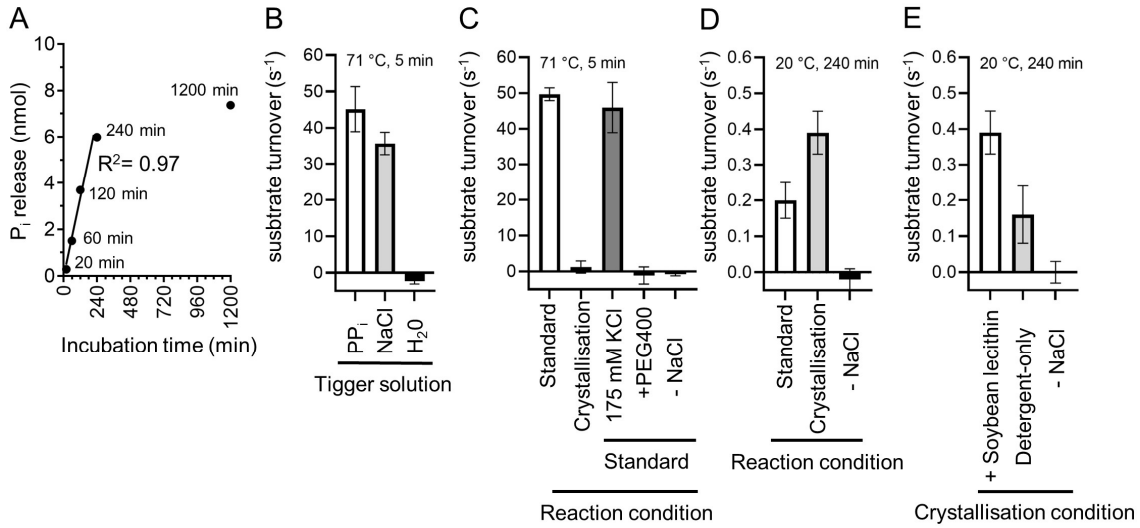


Figure 6.5: Quantitative P_i-release activity assays of *Tm*-PPase in a range of different reaction conditions. (A) Assays optimisation to ensure P_i release by *Tm*-PPase is linear with time during incubation at 20 °C. Linear regression and coefficient of determination (R^2) shown. (B-C) Substrate turnover (s^{-1}) by *Tm*-PPase at 71 °C upon reaction initiation with different trigger solutions (B) or in a range of different reaction conditions (C). (D-E) Substrate turnover (s^{-1}) by *Tm*-PPase at 20 °C in different reaction conditions (D) or after different sample treatment (E). Negative controls lack NaCl in the final reaction condition (-NaCl).

Tm-PPase did not show any activity when assayed in "crystallisation-like" solution instead of standard reaction mixture at its optimal reaction temperature (71 °C) (Figure 6.5C). The "crystallisation-like" solution only differed in the KCl concentration, which was increased to 175 mM, and the addition of 26% *v/v* PEG400 to the standard condition (Table 6.2). In subsequent assays, PEG400 addition was identified as cause for the activity loss in *Tm*-PPase (Figure 6.5C). Intriguingly, protein was active in "crystallisation-like" solution and even exceeded activity levels obtained in standard reaction mixture when the assay was performed at moderate temperatures (20° C) that are suitable for time-resolved crystallographic studies (Figure 6.5D). Nevertheless, the substrate turnover was overall much slower at below-optimum temperatures with a maximum k_{cat} of 49.67 ±0.45 s^{-1} at 71 °C and a k_{cat} of 0.39 ±0.08 s^{-1} at 20 °C.

For a better approximation of *in crystallo* reaction rates, the activity of non-relipidated protein (detergent-only) was determined. Without soybean lecithin relipidation, as in crystals, the substrate turnover further dropped to 0.16 ±0.09 s^{-1} (Figure 6.5E). This translates to a single PP_i molecule being turned over every ~6.25 seconds and is slow enough for manual crystal soaking and flash cooling to map intermediate catalytic states by X-ray crystallography. Actual *in crystallo* reaction rates could not be obtained as

crystals dissolved over time when transferred into PEG-supplemented standard reaction conditions, but might be even slower due to crystal contacts rigidifying the protein.

6.2.4. Time-resolved X-ray crystallography

The time-resolved X-ray crystallography experiments were guided by activity measurements using “crystallisation-like” conditions (6.2.3). The time window monitored focused on the first 10 seconds upon reaction initiation by Na^+ addition but was extended to include reaction times of up to 3600 seconds to account for increased rigidity of crystallised protein. Several vapour-diffusion crystals were soaked in Na^+-containing crystallisation solution for 1, 2, 5, 10, 60, 300, 600 and 3600 seconds to trigger the reaction, quickly flash cooled in liquid nitrogen to stop it and shot at P14-I (DESY) to resolve the enzyme state. Due to severe anisotropy, the diffraction data was submitted to the StarAniso webserver for anisotropic data correction and cut-off of merged intensity data at a local weighted average I/σ(I) of 1.2 (default parameter). *Tm*-PPase structures were solved by molecular replacement using the *Tm*-PPase:CaMg resting state structure (PDB: 4AV3) with hetero atoms removed as search model, which placed one *Tm*-PPase dimer in the asymmetric unit (Table 6.3-Table 6.4). Non-soaked reference crystals (t=0 seconds) diffracted to 2.65 Å along h, 3.32 Å along k and 3.79 Å along l. At the given resolution, the transmembrane helix backbone could be reliably placed into the $2F_o$-F_c electron density at a contour level of 1 σ allowing a good global overview and comparison to previously published M-PPase structures. Regions rich in low molecular weight residues such as glycine or proline, residues with flexible side chains such as arginine or lysine and most loop regions are only poorly defined or absent in the $2F_o$-F_c map, hampering the structure-function analysis.

The t=0 s reference structure was carefully refined considering electrostatic interactions, steric hinderance and geometry restraints. Average NCS maps and feature-enhanced maps further helped with modelling of poor electron density regions. In the absence of Na^+, *Tm*-PPase was conformationally trapped in a resting-state almost identical to the inhibited *Tm*-PPase:CaMg structure (rmsd/C_α: 0.41 Å). Subunit A and B were refined individually and were also almost identical with a rmsd/C_α of 0.21 Å. Despite the presence of PP_i in the crystallisation condition, it was not located at the active site, nor did it bind anywhere else. Most conserved aspartate and glutamate residues appeared flexible as indicated by the lack of $2F_o$-F_c density in the absence of ligands at the active site. Mg^{2+} or water was not resolved at the given resolution in this region.

Table 6.3: X-ray data collection and refinement statistics of time-resolved (0-10 s) *Tm*-PPase structures from the single best dataset.

Data collection	**t = 0 s**	**t = 1 s**	**t = 2 s**	**t = 5 s**	**t = 10 s**
Dataset	single best	single best	single best	single best	single best
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Cell dimensions					
a, *b*, *c* (Å)	84.0, 110.2, 108.0	84.0, 110.6, 107.6	83.7, 109.5, 106.8	84.0, 110.0, 107.5	84.1, 110.9, 107.2
α, β, γ (°)	90.0, 107.9, 90.0	90.0, 108.0, 90.0	90.0, 108.1, 90.0	90.0, 108.2, 90.0	90.0, 108.0, 90.0
Source	DESY: P14-I	DESY: P14-I	DESY: P14-I	DESY: P14-I	DESY: P14-I
Wavelength (Å)	0.967	0.967	0.967	0.967	0.967
Resolution (Å)	75.34-2.65 (3.08-2.65)	79.88-2.66 (2.76-2.66)	101.53-2.50 (2.86-2.50)	64.59-2.88 (3.21-2.88)	110.94-2.51 (2.88-2.51)
Overall (Å)	2.65	2.66	2.50	2.88	2.51
along h axis	2.65	2.66	2.50	2.88	2.51
along k axis	3.32	3.40	3.02	3.30	3.36
along l axis	3.79	3.54	3.71	4.03	3.362
Measured reflections	194305 (9787)	105617 (4804)	118912 (5793)	52824 (3022)	111217 (5636)
Unique reflections	27956 (1398)	29239 (1455)	33036 (1636)	20150 (999)	30634 (1529)
Completeness (%)	91.6	91.3	92.3	75.3	92.0
$CC_{1/2}$	0.999	0.997	0.998	0.995	0.997
Mean I/σ(I)	12.3	7.9	9.1	9.3	9.2
Multiplicity	7.0	3.6	3.6	2.6	3.6
B-factors ($Å^2$)	108.65	69.09	78.43	103.63	69.10
R_{merge}	0.064 (0.993)	0.082 (0.611)	0.056 (0.539)	0.042 (0.588)	0.068 (0.603)
R_{meas}	0.069 (1.073)	0.096 (0.725)	0.065 (0.636)	0.052 (0.711)	0.079 (0.076)
R_{pim}	0.026 (0.042)	0.050 (0.385)	0.034 (0.333)	0.030 (0.395)	0.041 (0.468)
Refinement	**t = 0 s**	**t = 1 s**	**t = 2 s**	**t = 5 s**	**t = 10 s**
Molecular replacement search model	*Tm*-PPase: CaMg	*Tm*-PPase: CaMg	*Tm*-PPase: CaMg	*Tm*-PPase: CaMg	*Tm*-PPase: CaMg
Active site modelled	empty	empty	empty	empty	empty
R_{work}(%)/R_{free}(%)	23.8/27.4	26.8/28.0	22.9/24.0	25.5/28.7	23.76/25.22
No. of atoms	10330	9965	9965	9965	9965
Protein	10311	9965	9965	9965	9965
Ligands/Lipids	0	0	0	0	0
Water	13	0	0	0	0
No. of chains (ASU)	2	2	2	2	2
B-factors ($Å^2$)	114.96	72.41	88.17	103.83	74.30
Protein	114.99.	72.41	88.17	103.83	74.30
Ligands/Ions	103.14	-	-	-	-
R. M. S. Deviations					
Bond lengths (Å)	0.001	0.003	0.004	0.003	0.003
Bond angle (°)	0.35	0.63	0.70	0.63	0.70
Ramachandran statistics (%)					
Favoured	97.86	94.44	94.37	94.87	94.30
Allowed	2.14	4.77	4.77	4.33	4.69
Outliers	0.00	0.79	0.87	0.63	1.01

Statistics for the highest-resolution shell are shown in parentheses

Table 6.4: X-ray data collection and refinement statistics of time-resolved (60-3600 s) *Tm*-PPase structures from the single best dataset.

Data collection	**t = 60 s**	**t = 300 s**	**t = 600 s**	**t = 3600 s**
Dataset	single best	single best	single best	single best
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Cell dimensions				
a, *b*, *c* (Å)	84.1, 109.7, 107.8	84.1, 111.1, 106.2	83.5, 111.0, 106.1	83.3, 111.1, 106,1
α, β, γ (°)	90.0, 107.8, 90	90.0, 108.2, 90.0	90, 109.1, 90	90.0, 109.0, 90.0
Source	DESY: P14-I	DESY: P14-I	DESY: P14-I	DESY: P14-I
Wavelength (Å)	0.967	0.967	0.967	0.967
Resolution (Å)	109.68-2.84 (3.23-2.84)	100.88-3.59 (4.07-3.59)	78.93-3.68 (3.983-3.68)	100.37-4.81 (5.56 – 4.81)
Overall (Å)	2.84	3.59	3.68	4.81
along h axis	2.84	3.59	3.75	4.81
along k axis	3.40	4.12	3.68	5.67
along l axis	3.83	4.49	4.59	6.22
Measured reflections	90910 (4045)	50005 (3962)	52549 (2619)	18934 (1006)
Unique reflections	25374 (1261)	14362 (1105)	14505 (725)	5220 (260)
Completeness (%)	90.0	90.7	90.0	85.1
$CC_{1/2}$	0.990	0.998	1.00	0.999
Mean I/σ(I)	5.3	3.9	12.4	12.2
Multiplicity	3.6	3.5	3.6	3.6
B-factors (Å^2)	103.72	146.15	215.73	294.54
R_{merge}	0.112 (0.447)	0.092 (0.766)	0.031 (1.342)	0.042 (1.013)
R_{meas}	0.132 (0.532)	0.109 (0.901)	0.037 (1.576)	0.050 (1.176)
R_{pim}	0.070 (0.285)	0.057 (0.471)	0.019 (1.039)	0.026 (0.586)
Refinement	**t = 60 s**	**t = 300 s**	**t = 600 s**	**t = 3600 s**
Molecular replacement search model	*Tm*-PPase:CaMg	*Tm*-PPase:CaMg	*Tm*-PPase:CaMg	*Tm*-PPase:CaMg
Active site modelled	empty	empty	empty	Empty
R_{work}(%)/R_{free}(%)	23.7/25.0	25.72/29.88	30.38/32.26	30.33/35.76
No. of atoms	9965	9965	9965	9965
Protein	9965	9965	9965	9965
Ligands/Lipids	0	9	9	0
Water	0	0	0	0
No. of chains (ASU)	2	2	2	2
B-factors (Å^2)	101.32	161.17	200.25	356.84
Protein	101.32	161.17	200.25	356.83
Ligands/Ions	-	-	-	-
R. M. S. Deviations				
Bond lengths (Å)	0.006	0.003	0.003	0.007
Bond angle (°)	0.65	0.62	0.64	1.04
Ramachandran statistics (%)				
Favoured	94.73	94.95	95.38	94.01
Allowed	4.40	4.19	3.90	4.91
Outliers	0.87	0.87	0.72	1.08

Statistics for the highest-resolution shell are shown in parentheses

The diffraction quality of crystals that were soaked in trigger solution for up to 60 seconds was similar to non-soaked reference crystals (t=0 s). After 300 seconds reaction initiation, the diffraction quality declined abruptly (Figure 6.6A). Reflections were detectable up to 3.59 Å after 300 seconds, up to 3.68 Å after 600 seconds and up to 4.81 Å after 3600 seconds in the best direction. Interestingly, the diffraction quality did not decline linearly with time but was linked to the active site status. Up to 300 seconds soaking, the active sites of both subunits remained empty and side chains unordered. After 600 seconds of soaking, positive F_o-F_c density appeared at 3 σ in active site of subunit A, but not in subunit B (Figure 6.6B). This is the first time an asymmetric occupation of the active site was observed in a M-PPase structure. After 3600 seconds of soaking, positive F_o-F_c density appeared in subunit B as well, which demonstrates unrestricted access to both sites and shows that asymmetric binding is not a crystallographic artefact. Unfortunately, the $2F_o$-F_c side chain density becomes intractable after 300 seconds of soaking, which hampers the analysis of local conformational changes at regions of interest such as the active site and ion gate. However, comparison of the Cα transmembrane helix backbone to other conformational states shows that the overall structure remains very similar to the resting-state *Tm*-PPase:CaMg structure with an rmsd/C_α between 0.41-0.48 Å for all time-points captured.

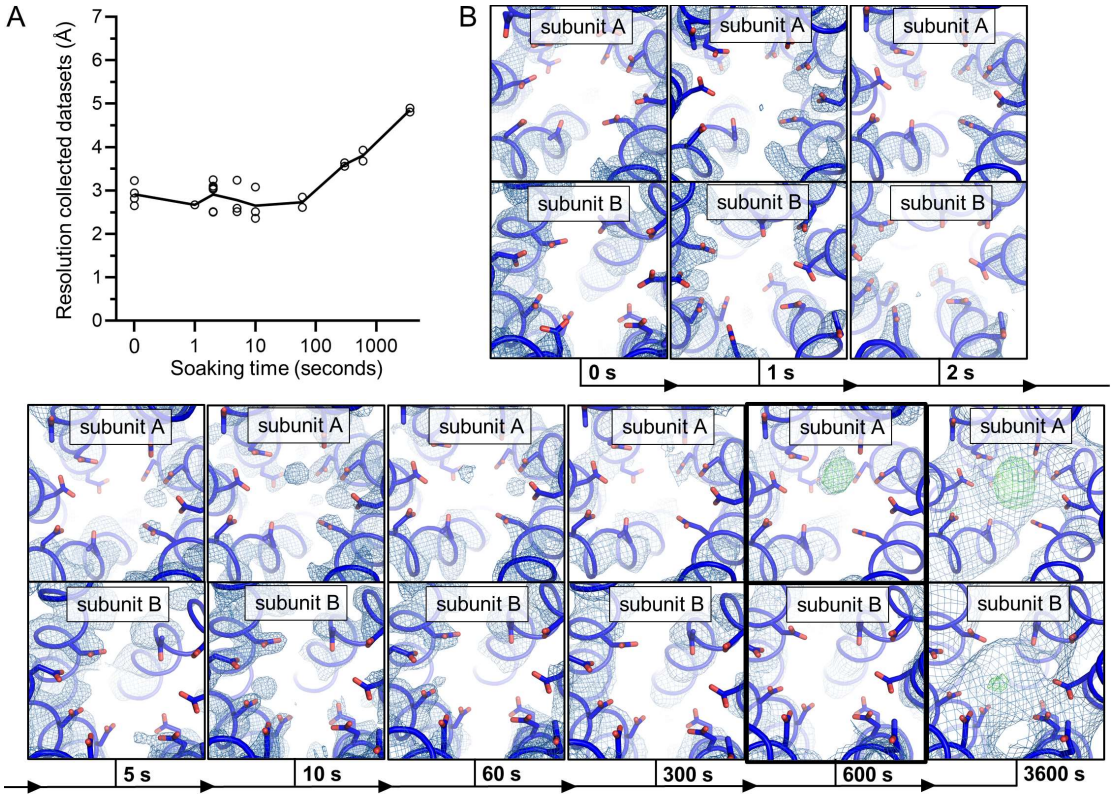


Figure 6.6: Comparison and structural overview of time-resolved *Tm*-PPase structures. (A) Diffraction quality at different time-points. Each collected dataset is represented by a circle with the diffraction in the best direction plotted. The mean resolutions of each time point are connected by a black line. (B) *Tm*-PPase active site of subunit A and subunit B with $2F_o$-F_c density (blue) and F_o-F_c density (red/green) shown. The t=600 seconds dataset is highlighted by a black box and shows occupation of the active site in subunit A but not subunit B. If not stated otherwise, the $2F_o$-F_c density is contoured at 1 σ and the F_o-F_c density is contoured at 3 σ.

For time reasons and due to the absence of significant structural changes, only the reference structure in the absence of Na^+ was fully refined. To improve the electron density maps, datasets were combined into 23 different clusters after isomorphism analysis based on unit cell variation in BLEND (CCP4i) (Figure 6.7A). BLEND assesses isomorphism by means of the linear cell variability (LCV), which quantifies the change of the three unit cell axes as percentage (281). Values below 2% typically indicate a fairly good isomorphism between datasets as structural changes only become noticeable at >1.5% (281). There is a distinct change in the unit cell dimensions of data collected after 0-300 seconds reaction initiation (cluster A) and data collected after 600-3600 seconds reaction initiation (cluster D). This coincides with the drop of resolution and change of the active site occupation status. The combination of all collected data (Cluster A+D) results in an LCV of 4.1%, whereas the individual clusters A and D have an LCV of 2.8% and 0.5%, respectively. No pattern regarding the soaking time emerges when these clusters are further subdivided. For example, cluster A can be subdivided into cluster B and cluster C, which both include data from 0-300 seconds and 2-300 seconds but have a lower LCV of 2.0% and 1.4%, respectively. Several data quality parameters such as the multiplicity, completeness, mean I/σ(I) and maximal resolution improved in combined datasets while R_{pim} remained stable and within the generally accepted limit of ~5% (Table 6.5). This also translates to the electron density map quality, which showed the most improvement when data of cluster A was used even though the LCV was slightly above 2%, which is not a hard limit for isomorphism. Therefore, cluster A was chosen for downstream processing, structure solution and careful refinement (Table 6.5).

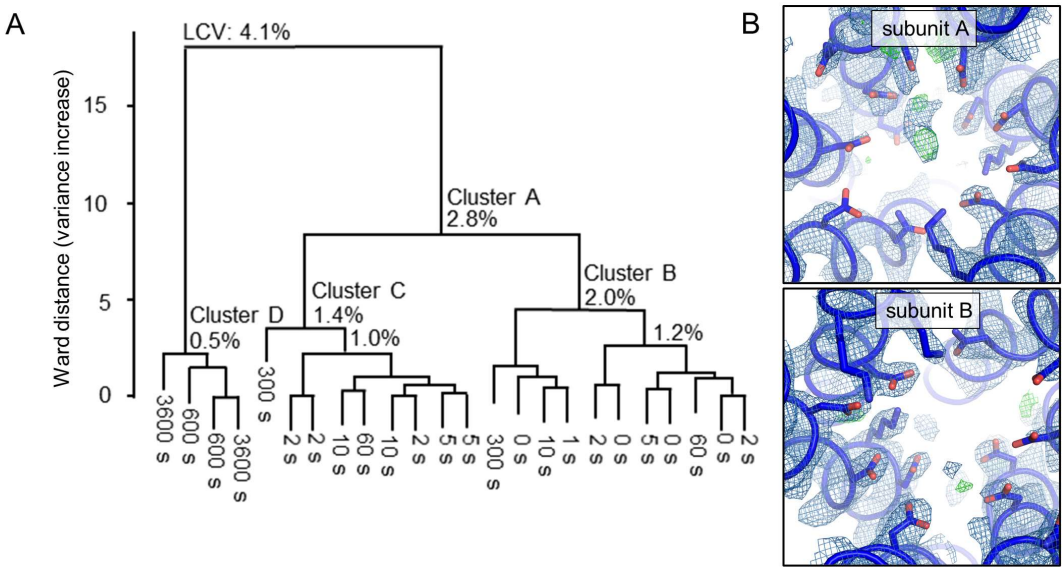


Figure 6.7: Comparison and structural analysis of combined datasets. (A) Dendrogram of BLEND analysis to identify isomorphous time-resolved datasets. Nodes of the four biggest cluster are labelled with the linear cell variability (LCV). (B) *Tm*-PPase active site of Cluster A (t = 0-300s) shown with $2F_o$-F_c density (blue) and F_o-F_c density (red/green). If not stated otherwise, the $2F_o$-F_c density is contoured at 1 σ and the F_o-F_c density is contoured at 3 σ.

Table 6.5: Comparison of individual and combined *Tm*-PPase datasets (clusters)

Datasets	single best	Cluster A	Cluster B	Cluster C
Dataset	t = 0 s	t = 0-300s	t = 0-300s	t = 2 -300 s
Maximal resolution (Å)	75.34-2.65	75.19-2.52	102.23-2.52	75.04-2.52
Completeness (%)	91.6	94.3	94.2	93.9
$CC_{1/2}$	0.999	0.995	0.996	0.988
Mean I/σ(I)	12.3	13.1	11.7	9.4
Multiplicity	7.0	55.2	39.6	22.1
R_{pim}	0.026	0.036	0.035	0.044

I could reliably model most side chains, including functionally important lysine such as $K^{16.50}$ at the ion gate, place detergent alkyl chains at the dimer interface as seen in *Vr*-PPase:Mg_5IDP, add several structural water molecules and build additional loop regions, for example $loop_{5\text{-}6}$, using data combined in cluster A. During refinement, R-factors ($R_{work/free}$) improved from 23.8%/27.4% for the best individual dataset (t=0 s) to 22.5%/24.7% (t=0-300 s) and the average B-factor dropped from 114.96 $Å^2$ to 75.71 $Å^2$, respectively. Overall, the new model remained very similar to the resting state *Tm*-PPase:CaMg structure with a rmsd/C_α of 0.31 Å. In contrast to individual datasets that were collected after 0-300 seconds reaction initiation, the active site of combined data showed $2F_o$-F_c and positive F_o-F_c density (Figure 6.7B). A range of hetero atoms (Mg^{2+}, H_20, PP_i, P_i, glycerol or Tris) were added, but none fitted the density well even when adjusting occupancy levels (Figure S8, Appendix). Due to the time-resolved nature of the experiments the active site of datasets from different time-points might be differently occupied, which would not be detectible by LCV analysis if no or only local residue reorientations occur in the monitored time window. Therefore, the new model derived from cluster A was used as starting point to refine clusters that only combined datasets from identical time-points (Table S11-Table S12, Appendix). For time reasons I mainly focused on the analysis of the active site and ion gate occupation in subunit A and subunit B.

In general, the data quality of combined datasets that were collected at the same timepoint was superior to individual datasets (Table S11-Table S12, Appendix). However, distinct structural changes between the individual and combined datasets were only visible for the 300-seconds structure at the active site. In contrast to the individual datasets, $2F_o$-F_c and positive F_o-F_c density appeared in subunit A at 1 σ and 3 σ, respectively. The density was fit best by P_i with an occupancy of 100% or PP_i with an occupancy of 50% (Figure 6.8A, Figure S9 Appendix), but only the latter also fitted to local B-factor distribution (Table 6.6). After 600 seconds of soaking, the density and local B-factor distribution was fit best when the occupancy of PP_i at the active site of subunit A was increased to 100%. However, it appears that PP_i did not arrive yet in its final binding pose as it bound 3 Å above the IDP-binding site in *Tm*/*Vr*-PPase:Mg_5IDP

structures (Figure S10, Appendix). After 3600 seconds of soaking, both subunits displayed $2F_o$-F_c and positive F_o-F_c density at the active site as in all IDP-bound structures. Unfortunately, the resolution was too bad to reliably place PP_i or P_i at either active site at this timepoint (Figure 6.8A, Figure S9, Appendix).

Table 6.6: Local B-factor distribution at active site of time-resolved *Tm*-PPase structures from different time-points

Soaking time (s)	Dataset	Subunit	Active site	Occupancy	B-factor*	ΔB-factor†
300	combined	A	residues∓	1	98	-
			PP_i	1	155	57
			PP_i	0.5	114	16
			P_i	1	143	45
600	combined	A	residues∓	1	233	-
			PP_i	1	248	15
			PP_i	0.5	218	15
			P_i	1	250	17
3600	combined	A	residues∓	1	373	-
			PP_i	1	384	11
			P_{i2}	1	Pi_1: 356 Pi_2: 350	17 23
			P_i	1	351	22
	combined	B	residues∓	1	389	-
			PP_i	1	419	30
			P_{i2}	1	Pi_1: 402 Pi_2: 378	13 20
			P_i	1	384	5

∓ Selection of local Cα atoms within a 5 Å of positive density at the active site
* Isotropic B-factor model. B-factor of side chain and main chain grouped in refinement.
† Absolute difference to B-factor of active site residues nearby

Although the active site occupation changed over time, helix orientations remained highly symmetrical between both subunits and overall similar to the resting-state *Tm*-PPase:CaMg structure with a rmsd/C_α of 0.33 Å and 0.48 Å, respectively. Consequently, there were also no major structural changes observed at the ion gate at time-points (up to 600 seconds post reaction initiation) that allowed the unambiguous placement (*e.g.* no frameshifts) of helices into $2F_o$-F_c density (Figure 6.8B). However, there is a change at the semi-conserved glutamate. $E^{6.53}$ appears to be flexible in low Na^+ conditions as indicated by the absence of $2F_o$-F_c density or negative density when modelled as seen in *Tm*-PPase:CaMg (Figure 6.8B). All other side chain orientations at the ion gate were well defined in the $2F_o$-F_c map up to 300 seconds of soaking. Typically, the semi-conserved glutamates is well ordered and side chain density visible at much lower resolutions. Its flexibility in low Na^+ conditions may have implications for the ion-selectivity at sub-physiological Na^+ concentrations in K^+-dependent Na^+-PPases and is discussed in 6.3.4.

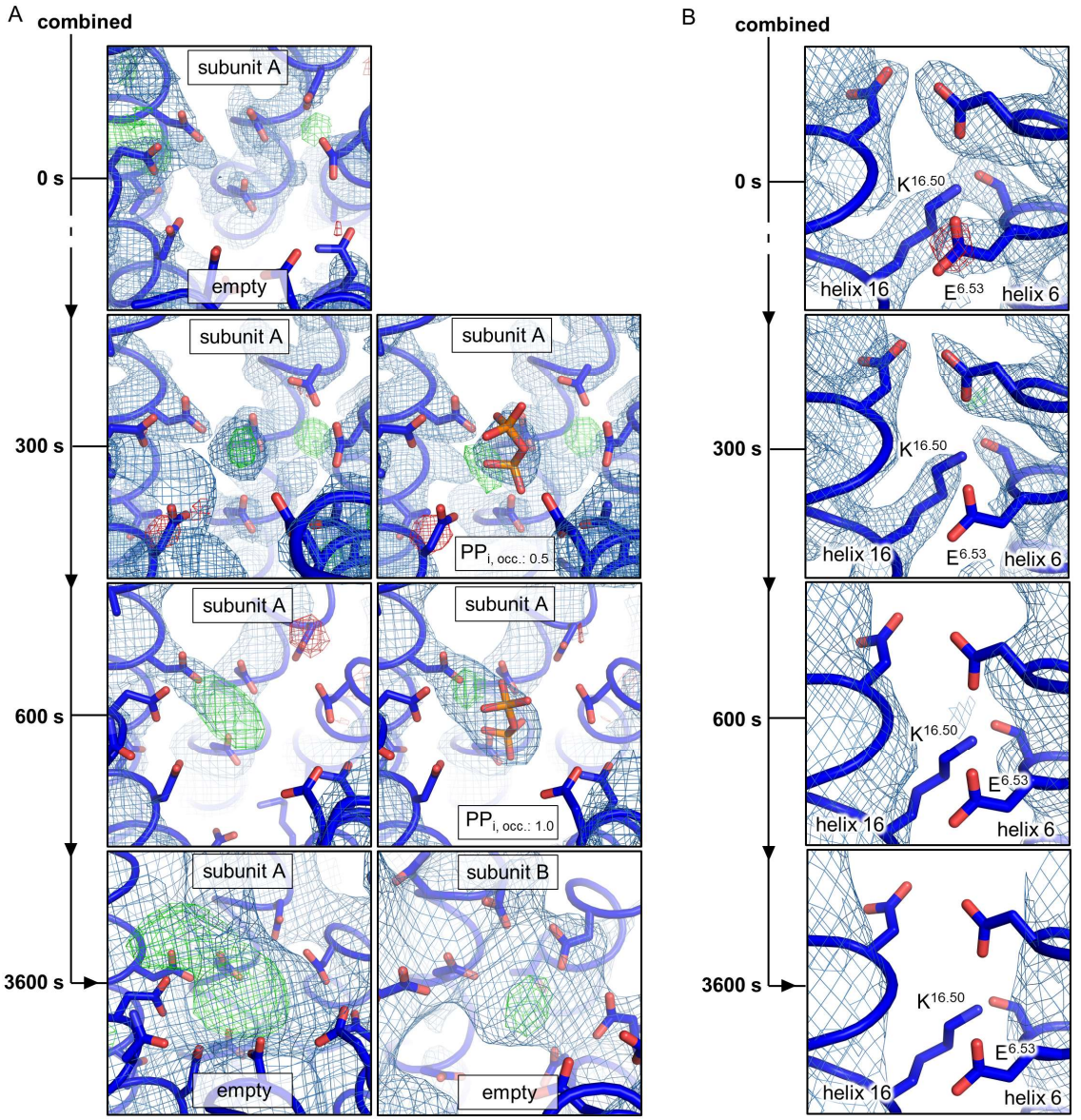


Figure 6.8: Enzymatic core regions of time-resolved *Tm*-PPase structures obtained from combined datasets. Active site (A) and ion-gate (B) at t= 0 s, 300 s, 600 s and 3600 s shown with $2F_o$-F_c density (blue) and F_o-F_c density (red/green). Where hetero atoms were modelled into the active site the resulting F_o-F_c (red/green) after refinement is displayed. The ligand identity and occupancy (occ) are displayed in the lower section of each panel. If not stated otherwise, the $2F_o$-F_c density is contoured at 1 σ and the F_o-F_c density is contoured at 3 σ. PP_i/P_i orientations modelled with guidance of *Tm*/*Vr*-PPase:Mg_5IDP structures.

6.3. Discussion

The use of ligands including inhibitors for conformational stabilisation/trapping is an valuable tool for mapping different catalytic states and proved key for the structural characterisation of challenging targets such as GPCRs (282). Similarly, structural studies of M-PPases heavily relied on ligand binding to map different catalytic states in the past (127–129,131). However, static snapshots of protein states can only provide limited insight into mechanistic details of biological processes and may even be misleading when interpreting structure-function relationships without considering protein motion, the driving force of said biological processes. Therefore, the structural biology field is shifting towards the view that dynamics, structure and function are of equal importance for an in

depth understanding of protein biology (275). Here, I investigated asymmetric *Tm*-PPase conformations in time-resolved structural studies without using inhibitors to avoid imposing symmetry on M-PPase subunits (131).

6.3.1. Feasibility of rapid mixing and manual crystal handling

Vapour-diffusion crystallisation, and later batch crystallisation, were based on a condition that yielded the *Tm*-PPase:Mg_5IPD structure and optimised for a time-resolved approach by trapping *Tm*-PPase in the absence of Na^+, which is essential for its activity as K^+-dependent Na^+-PPases (129). Crystals were additionally optimised to grow as separate 3D objects with sharp edges. Although crystal morphology cannot be taken as a reliable indicator of mosaicity or diffraction quality it might give an indication of twinning. This is best avoided as it causes issues in downstream data processing due to the overlapping or superposition of reflection spots (283). Optimised crystals were of different unit cell geometry (monoclinic, $P2_1$) compared to crystals yielding the *Tm*-PPase:Mg_5IPD structure (orthorhombic, $P2_12_12_1$) and better diffraction quality despite not using any inhibitors for conformational stabilisation (129). Unfortunately, the improved crystal characteristics could not be translated into batch crystals even when using identical final crystallisation conditions, which shows that not only the end point in a crystallisation phase diagrams is important but also how to get there. Here, vapour-diffusion and batch crystallisation conditions were different during nuclei formation (Figure 6.9). The resulting differences in crystal geometry and diffraction quality may thus be overcome by micro seeding batch crystallisation trials with vapour diffusion crystal seeds. High quality batch crystallisation has the distinct advantage to be compatible with time-resolved experiments on HARE chips that can be used to map biological processes completed within milliseconds (192).

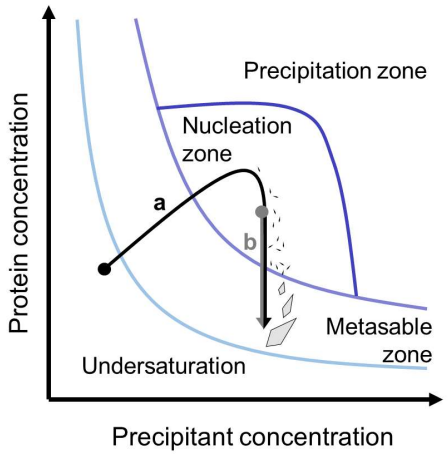


Figure 6.9: Schematic crystallisation phase diagram. *Tm*-PPase vapour diffusion (black line) and batch (grey line) crystallisation paths are shown. The boundary between undersaturation, the metastable zone, the nucleation zone and the precipitation zone are indicated by blue lines. Drop equilibration in vapour-diffusion is represented by moving diagonally across the phase diagram (a). During nuclei formation in the nucleation zone the protein concentration in solution drops, which is represented by the vertical move across the phase diagram (b). This continues in the metastable zone where nuclei from both approaches grow into 3D crystals.

In case of *Tm*-PPase, time-resolved structural studies only require a time resolution in the seconds time scale as shown by activity assays that resembled “crystallisation-like” conditions. At 20 °C, substrate turnover in *Tm*-PPase was slow enough for manual reaction initiation by rapid mixing (soaking) of vapour-diffusion crystals with trigger

solution and subsequent flash cooling in liquid nitrogen to stop the reaction. Recently, a similar approach proved successful to follow the replicative DNA synthesis by a bacterial polymerase (284). The authors of this study slowed down the primer extension by lowering the Mg^{2+} concentration and then solved time-resolved DNA polymerase structures (60 seconds – 25 hours) after reaction initiation in nucleotide containing trigger solution and subsequent manual flash cooling in liquid nitrogen to stop DNA replication (284). In line with the observations in *Tm*-PPase, the obtained time-resolved structures in DNA polymerase only differed at the active site during the monitored time period (284). Unfortunately, the diffraction quality of membrane protein crystals is typically considerably worse compared to crystals of soluble proteins, as was the case for *Tm*-PPase, which hindered a comparable analysis of the active site configuration here.

6.3.2. Reliability of activity data from "crystallisation-like" conditions

As previously discussed, the quantitative P_i-release assay performs poorly at high temperatures (>40 °C) due to issues of Mg_2PP_i precipitation (4.3.6). Therefore, substrate turnover rates obtained at 71 °C may be underestimated. Moreover, data suggesting PEG inhibition at 71 °C must also be interpreted with care as it cannot be distinguished between pseudo inhibitory effects caused by PEG-induced Mg_2PP_i precipitation and actual inhibitory effects caused by thermal decomposition products of PEG, for example low molecular weight esters (285). PEG itself did not inhibit *Tm*-PPase as it enhanced protein activity at 20 °C. This may be explained by its "crowding effect", which leads to increased effective concentrations of co-factors or substrate and possibly more effective binding to protein thereof (286,287). Similar effects of crowding agents on protein activity have been observed for DNA ligase and DNA polymerase (286,287). Alternatively, specific interactions of PEG with *Tm*-PPase could alter protein activity, but no PEG binding was observed in any crystal structures so far.

All data relevant for the design of the time-resolved X-ray crystallographic experiment were collected at 20 °C in "crystallisation-like" conditions and are therefore not affected by issues of Mg_2PP_i precipitation (personal communication with Dr. Baykov). Nevertheless, the true *in crystallo* reaction rates may be much slower than estimated based on data obtained in solution (k_{cat}: 0.16 s^{-1}) due to the rigidity imposed by the crystal onto individual protein molecules (284). Indeed, the crystallographic data showed that the reaction cycle was not complete within 6.25 seconds. Therefore, the monitored time period in for time-resolved X-ray crystallographic experiments was increased to 3600 seconds.

6.3.3. Asymmetric substrate binding to active site

This is the first report of time-resolved M-PPase structures and the first time that binding of the physiological substrate (PP_i) to the active site has been resolved. Most importantly, substrate binding to the active site was asymmetric. This supports the idea of functional asymmetry and a „half-of-the-sites" reactivity model in M-PPases (131,136). As expected and seen in other time-resolved structural studies, the occupancy of substrate at the active site increased with time (284). However, I was not able to record structural rearrangements upon binding of the substrate as the resolution drops precisely when PP_i enters the active site. This may be a consequence of crystal contact formation at the active site and indicative of structural changes (Figure S11, Appendix), but hampers to pinpoint the latter. However, the data quality remains sufficient to distinguish between helical orientations such as the downward movement of helix 12 and the corkscrew motion of helix 6 and 16 that are characteristic for resting-state and product-bound protein conformations. Therefore, major conformational changes did not take place in the monitored time period, which is in line with the observation that PP_i did not arrive at its proper binding position yet. It is unlikely that the mapped PP_i position represents the true orientation of physiologically relevant substrate in M-PPases as this would invalidate the models of enzyme activation (129) and K^+-independence (136) that are coherent with all other structural and functional data published so far.

Before the arrival of PP_i at the active site, the data quality is of sufficient quality for a more detailed structure function analysis. In absence of Na^+, *Tm*-PPase crystallises in the resting-state, which has been previously mapped by using inhibitory Ca^{2+} instead of withdrawing Na^+ (127). Both Ca^{2+} and the withdrawal of Na^+ hinder PP_i binding to the active site, which explains why they stabilise the same conformational state. Ca^{2+} replaces Mg^{2+}, which is required for the formation of a metal cage that binds PP_i (127), whereas Na^+ prepares the active site for the arrival of PP_i and needs to enter first as there is not enough space to pass through once the active site is occupied (129).

6.3.4. Ion selectivity at low Na^+ concentrations

The resting-state has been mapped previously, but there is a distinct structural change at the ion gate compared to the published structure that may be a result of the absence of Na^+. The semi-conserved glutamate, $E^{6.53}$ in *Tm*-PPase, is not hold in place by $K^{16.50}$ but appears to be flexible. All other neighbouring residues, including lysine, are resolved in the electron density and thus well ordered. This has mechanistic implications for the ion selectivity at low Na^+ concentrations and can explain why K^+-dependent Na^+-PPases are also capable to transport H^+ at sub physiological Na^+ concentrations (259) in the context of the updated model of ion selectivity (4.3.5). As $E^{6.53}$ is not tightly locked in a position to coordinate the ε-NH_3^+ group of $K^{16.50}$, it can adapt an orientation that

resembles the ion-gate set up in K^+-independent H^+-PPases (4.3.5). This would facilitate its protonation to account for the negative charge in a highly hydrophobic environment and consequently lead to H^+-pumping.

6.4. Future directions

In summary, this study is proof of concept for time-resolved X-ray crystallographic studies with M-PPases and paves the way for the elucidation of physiologically more relevant conformations including intermediate states in these enzymes. Although the loss of resolution upon PP_i arrival and the coarse timepoint sampling was not sufficient to elucidate mechanistic insights into functional asymmetry, the presented preliminary data already contributes to a more complete understanding of ion-selectivity and provides additional evidence for functional asymmetry and a „half-of-the-sites“ reactivity model (Figure 1.20) in M-PPase catalysis.

I followed a simplified approach to time-resolved crystallographic studies of M-PPases that does not require high-density crystal suspensions, which is a major limiting factor, especially when working with membrane proteins (275). However, poor X-ray diffraction upon active site occupation hampered the dynamics-structure-function analysis. To improve the data quality in future experiments, more single crystals can be tested and datasets from the same time-points merged. This approach already proved suitable to improve the electron density maps of the time-resolved resting state structure obtained from data collected after 0-300 seconds reaction initiation. Further improvements to the crystal quality by additional optimisation of crystallisation conditions might be difficult within the limits set by the strict requirements of M-PPases for enzymatic activity. Moreover, the loss of resolution upon PP_i arrival is likely caused by crystal contact formation at the active site of subunit B and can only be resolved by changing the crystal system, which is not trivial. Therefore, time-resolved structural studies by cryo-EM are a promising alternative to X-ray crystallography. With the availability of a new generation of electron detectors of unmatched speed and sensitivity (288) and continued advances in automated image processing (289,290) near-atomic resolutions are no longer a rarity reserved to well behaved and highly symmetrical targets (52). Microfluidic spraying devices enable fast dispense-to-plunge times within milliseconds for grid preparation (291). This is fast enough to capture intermediate states in M-PPases in solution, even when working with mesophilic proteins at ambient temperatures such as *Cl*-PPase.

Chapter 7. Key findings, overall conclusions and future directions

7.1. Key findings of the project

This PhD project is embedded in the RAMP-ITN (https://ramp-itn.eu/) and had two overall objectives as such. On the one hand, it aimed at providing M-PPase test samples for the network as to drive the development of general-purpose tools for membrane protein research including crystallisation. On the other hand, it addressed mechanistic questions of M-PPase biochemistry by using conventional and more advanced crystallographic approaches, which benefits structure-guided drug discovery efforts (131,168).

In the following, the key findings of sub-projects that were presented in Chapter 3-6 are summarised and linked to their impact in the field of membrane protein and M-PPase research later.

7.1.1. IMPROvER: The Integral Membrane Protein Stability Selector

One of the two tools that were developed as part of my PhD project to tackle issues of membrane protein instability is IMPROvER, a computational prediction pipeline of thermostabilising point mutations in membrane proteins (3.2). In contrast to other tools aiding the thermostabilisation of membrane proteins, IMPROvER works on a broad range of targets, is not limited by their size and does not require *a priori* structural information or naturally thermostable orthologues (3.3.1). I developed a rapid screening strategy for the mutagenesis and variant evaluation and used it to assess IMPROvER's performance on *Cl*-PPase, a K^+-dependent Na^+,H^+-PPase (3.2.1). The results of the *Cl*-PPase stabilisation campaign have formed the basis for the streamlined stabilisation of other membrane proteins that were investigated by colleagues, namely hENT and PTH_1R (173). By fine-tuning the strategy of variant selection from the prediction pool throughout the different protein stabilisation campaigns, IMPROvER reached unmatched success rates of up to 40% (PTH_1R) for the prediction of stabilising point mutations (3.3.1). The success rate of the *Cl*-PPase stabilisation campaign was 29%, which still outperforms comparable tools such as CompoMug (3.3.1). The best point-mutations in *Cl*-PPase, G130A and F20Y, increased the T_m by up to 9-13.0 °C while retaining wild-type like activity levels (3.2.1). As a result, these variants could be purified at higher concentrations (5.2.2) and in shorter-chain detergents (5.2.3) than wild-type protein, which both aids the crystallisation in LCP or by the vapour-diffusion method. The *Cl*-PPase stabilisation campaign demonstrates the potential of IMPROvER to aid the structural characterisation of inherently difficult-to-work with membrane proteins (7). In

future, IMPROvER will be made accessible *via* a webserver to simplify access and promote its broad application in the field of membrane protein research (3.4).

7.1.2. A novel high-throughput lipid screen for stabilising membrane proteins

The second tool that was developed to assist membrane protein stabilisation is the novel multi-purpose and high-throughput RAMP lipid screen (3.2). This pre-dispensed lipid screen allows the rapid identification of stabilising lipids with minimal effort and at affordable cost to reinstate physiologically relevant protein-lipid interactions in detergent-solubilised sample. It comprises 31 different lipids and lipid mixtures that covers a broad lipid space (3.1.2). I drove the screen development and evaluated its performance on three detergent-solubilised membrane proteins of different fold and mode of action ($A_{2A}R$, UapA, *Tm*-PPase) to demonstrate its general applicability and potential to facilitate membrane protein research (3.1.2). All three targets were stabilised by adding back lipids. The biggest effect on protein stability was observed upon relipidation of $A_{2A}R$ with CHS, which increased the T_m by +10.8 °C (3.1.2). This study did not only confirm well-established protein-lipid interactions but also revealed previously unknown lipid binding that affected protein stability, for example POPA stabilising $A_{2A}R$ or PSs stabilising UapA (3.1.2), and illustrates the value of the RAMP lipid screen for membrane protein research. The screen complements my effort to aid the functional and structural characterisation of highly challenging targets and can be used in tandem with IMPROvER to further boost membrane protein stability.

7.1.3. Structural basis for K^+-independence in M-PPases

As part of this project, I also presented the first K^+-independent H^+-PPase structure (*Pa*-PPase:Mg_5IDP) upon re-processing, extensive manual modelling and refinement of an old low-quality dataset (4.2.3). Most residues of enzymatic core regions, such as the active site and ion-gate, were well ordered and/or could be reliably placed when taking into account geometry restrains and electrostatic interactions (4.3.2). This enabled the structure-function analysis of the evolutionary conserved residue changes $A^{12.46}K$ and $A/G^{12.49}T$, which define K^+-independence in M-PPases. Only the former change was unambiguously linked to K^+-independence as ε-NH^{3+} of $K^{12.46}$ occupies a K^+-binding site and takes over substrate coordination (4.3.34.2). At the same time, this finding raises the question about the functional role of the coupled change at position 12.49. The introduction of an additional hydrogen bond in K^+-independent H^+-PPases between $T^{12.49}$ and $D^{11.50}$ may rigidify helix 12 movement against helix 11 (4.3.3). However, the biological implication of this remains unknown. The investigation of its putative involvement in substrate inhibition and functional asymmetry requires optimised and robust M-PPase activity assays for functional analysis (4.4), which are currently

being establishing by colleagues based on recent reports from expert M-PPase enzymologists (265).

7.1.4. An updated model of ion-selectivity in M-PPases

The K^+-independent H^+-PPase structure also revealed mechanistic details about ion-selectivity in M-PPases (4.3). The previous model of ion-selectivity was only valid for a small subset of M-PPase family members, namely K^+-dependent Na^+-PPases and K^+-dependent H^+-PPases, and failed when applied to K^+-independent H^+-PPases and Na^+,H^+-PPases (4.2.3.3). I identified structural changes at helix 5 that effect the residue orientation at the ion gate including $S^{5.43}$ and $E^{6.53/57}$, of which the latter (semi-conserved glutamate) has well-documented implications for ion-selectivity (4.2.3.3). These changes can explain how M-PPases can switch between H^+ and Na^+ pumping although the residue identity at the ion-gate is unchanged. I propose that the interplay of $S^{5.43}$ with $E^{6.53/57}$ is controlled by helix 5 orientation and dictates whether the negative charge of $E^{6.53/57}$ in the centre of the membrane spanning protein section is accounted for by protonation or Na^+-binding (4.3.5). In Na^+-PPases, the semi-conserved glutamate is forced to face away from $S^{5.43}$ to avoid clashing with $S^{5.43}$ upon substrate-binding induced helix reorientations and contributes to the formation of a Na^+-binding site (4.3.5). In H^+-PPases, $E^{6.53/57}$ is likely protonated as it does not contribute to the formation of a Na^+-binding site but hydrogen bonds to $S^{5.43}$ (4.3.5). This is made possible by the outward movement of helix 5 at the ion gate due to its tighter winding and straightening ($E^{6.53}$) or repositioning of the semi-conserved glutamate one helix turn down ($E^{6.57}$), which ultimately creates enough space for the interaction of $E^{6.53}$ with $S^{5.43}$ (4.3.5). The updated model of ion-selectivity is now valid for all M-PPase subclasses. It even holds when applied to K^+ dependent Na^+,H^+-PPases, for which no structural data is available yet, as ion pumping could simply be explained by asymmetric helix 5 orientations, and thus, a different $S^{5.43}$ $E^{6.53}$ interaction status between subunits (4.3.5). This idea is in line with growing evidence of functional asymmetry in M-PPase catalysis (131,136).

7.1.5. Asymmetric substrate binding demonstrating functional asymmetry in M-PPases

Following reports of functional asymmetry in M-PPase catalysis (131,136) and my findings that led to an updated model of ion-selectivity and suggest an asymmetrical *modus operandi* in K^+-dependent Na^+,H^+-PPase (7.1.4), I conducted time-resolved structural studies with the naturally thermostable model K^+-dependent Na^+-PPase from *T. maritima* (6.2). These experiments pave the way towards the structural characterisation of intermediate M-PPases conformations that are difficult or impossible to capture in conventional crystallographic approaches. In contrast to previous studies that relied on M-PPase inhibitors for conformational trapping/stabilisation, I took

advantage of the natural requirements for Na^+-PPase activity (6.1.2). *Tm*-PPase was crystallised in the absence of Na^+ but presence of PP_i and solved in a resting-state (6.2.2). M-PPase activity was then reinstated by addition of Na^+ to the activity assay reaction condition or crystallisation solution (6.2.4). The slow PP_i turnover at room-temperature allowed a manual time-resolved approach for X-ray crystallographic studies in which the reaction was triggered by crystal soaking in Na^+-containing crystallisation solution and stopped after a certain time-delay by flash cooling in liquid nitrogen (6.2.4). Following this simple strategy, I was able to map asymmetric substrate binding to the active site of *Tm*-PPase for the first time. This supports the „half-of-the-sites“ reactivity model that is based on kinetic data showing substrate inhibition and non-phosphorous inhibitors capturing asymmetric loop orientations (6.2.4). Unfortunately, substrate-binding to crystallised protein led to poor X-ray diffraction and thereby impaired a detailed dynamics-structure-function analysis (6.2.4). Nevertheless, this work is proof of concept for a straight-forward approach towards time-resolved structural studies of M-PPases that have the potential to yield biologically and pharmacologically interesting intermediate states. Our workgroup is particularly interested in mapping an intermediate gate-open structure to design state-specific inhibitors that would cause leaky membranes in M-PPase-carrying pathogens.

7.2. Future directions of M-PPase research

The findings of this PhD project answered as many questions as it led to new ones of which some are already being explored by our workgroup and collaborators. For example, what is the role of the evolutionary conserved secondary key change $A^{12.49}T$ of K^+-independence and the effect of the $T^{12.49}$-$D^{11.50}$ interaction in *Pa*-PPase? Is ion selectivity in K^+-dependent Na^+,H^+-PPase defined by asymmetric ion gate conformations? How do local residue configurations of enzymatic core regions and global helix orientations differ in asymmetric enzyme states and what are the biological consequences thereof?

Currently, experiments are ongoing to explore these questions in more detail. Dr. Antreas Kalli (University of Leeds) set up atomistic MD simulations (currently running) of wild-type and variant *Pa*-PPase to support the modelling of poorly defined key residues in the *Pa*-PPase:Mg_5IDP structure (*e.g.* $K^{12.46}$ and $K^{16.50}$) and explore the effect amino acid substitutions as position 12.49. This will be complemented by mutational studies investigating the effect of $S^{5.43}$ mutations on ion-selectivity and $T^{12.49}$ mutations on substrate inhibition in future. Moreover, Dr. James Hillier continued to work on the structural characterisation of *Cl*-PPase to validate the updated model of ion-selectivity in K^+-dependent Na^+,H^+-PPase. He also explores functional asymmetry and intermediate conformational states in time-resolved biophysical approaches. For this, a microfluidic

device for time-resolved Pulsed Electron-Electron Double Resonance (PELDOR) measurements is currently being developed by Dr. Christos Pliotas (University of Leeds) in collaboration with Prof. Arwen Pearson and co-workers (University of Hamburg).

In recent years cryo-EM has become a viable alternative to X-ray crystallography and targets even smaller than M-PPases are now in reach for routine near-atomic structure determination (52). This was mainly facilitated by the development of automated data processing pipelines (289,290) and highly sensitive direct-electron detectors (288), which together enable the compensation for beam-induced specimen motion from acquired movie frames and greatly improved the signal-to-noise ratio. As a result, less electron scattering events, which translates to smaller molecules, is sufficient for particle alignment and 3D reconstitution (293). In future, cryo-EM will play a bigger role for the structural characterisation of M-PPases. It does not require crystals, consumes less sample, and can yield several conformational states from a single dataset (294). The Astbury Biostructure Laboratory Centre of the University of Leeds is particularly well equipped for such studies as it has two state-of-the-art Titan Krios transmission electron microscopes. Moreover, time-resoled cryo-EM is being explored by Dr. Stephen Muench and co-workers, who are developing a micro-fluidic spraying device for time-resolved plunge-freezing of cryo-EM grids. This is already sufficiently fast to capture intermediate M-PPase conformations.

7.3. Significance of this work

The achievements, findings and impact of this PhD project are 3-fold.They comprise membrane protein research in general on the one hand and detailed mechanistic insights into M-PPase biochemistry on the other hand.

I contributed to the overall aim of “RAtionalising Membrane Protein crystallisation” of RAMP by suppling purified M-PPase sample for the method development of micro-fluidic dialysis chips at the Institute Laue-Langevin in Grenoble and contributed to, or led the development of, novel tools for protein stabilisation myself. Stable sample is a requirement for its biophysical characterisation and a major bottleneck for structural studies of membrane proteins (173). Previous approaches addressing issues of protein instability either relied on cost- and labour-intensive brute-force trial-and-error screening, which is often intractable for academic institutions, or were limited by the methodology itself, for example being only applicable to a small subset of targets or a certain target size (173). These downfalls were partly or fully addressed by the RAMP lipid screen and IMPROvER. Both tools have the potential to save resources that can be used more sensibly elsewhere, to accelerate the science of membrane proteins of which many have pharmacological relevance, and significantly improve protein robustness to simplify protein handling and improve protein yields.

I also contributed to a better understanding of the molecular mechanisms defining M-PPase biochemistry, particularly ion-selectivity. This aids our efforts to tweak M-PPase function or utilise M-PPases to fight existing and emerging challenges to human health and global food security. For example, their importance for the stress resistance in bacteria, parasites and plants gives leverage to address issues of multi-drug resistant pathogens or draught-induced crop loss (1.6). This will become even more important in future. By 2050 up to 183 million additional people will be at risk of hunger due to the global-warming induced global crop loss (UN, IPPC report 2019) and up to 50 million people are estimated to die from drug-resistant infections (UN, IACG report 2019) if no actions are taken.

Finally, I have made my research findings more accessible to the public in outreach events such as pub-talks, stalls at conferences or in a short movie. This is more important than ever in times of “fake-news”, deliberate disinformation campaigns and mistrust placed into science or scientific institutions. Scientists cannot hide in their laboratories anymore but must contribute to bridge the growing gap between the public and science. Otherwise, the loss of the public’s acceptance for a fact-based and science-guided approach to address current and future challenges of mankind, of which some are outlined above, would have catastrophic and potentially irreversible consequence for us and, more importantly, future generations.

References

1. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347,** (2015).
2. Rosenbaum, D. M., Rasmussen, S. G. F. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459,** 356–363 (2009).
3. Gadsby, D. C. Ion channels versus ion pumps: the principal difference, in principle. *Nat. Rev. Mol. Cell Biol.* **10,** 344–352 (2009).
4. Rees, D. C., Johnson, E. & Lewinson, O. ABC transporters: the power to change. *Nat. Rev. Mol. Cell Biol.* **10,** 218–227 (2009).
5. Dufrisne, M. B., Petrou, V. I., Clarke, O. B. & Mancia, F. Structural basis for catalysis at the membrane-water interface. *Biochim. Biophys. Acta* **1862,** 1368–1385 (2017).
6. Paulsson, M. Basement membrane proteins: structure, assembly, and cellular interactions. *Crit. Rev. Biochem. Mol Biol.* **27,** 93–127 (1992).
7. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discov.* **5,** 993–996 (2006).
8. Sriram, K. & Insel, P. A. G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs? *Mol. Pharmacol.* **93,** 251–258 (2018).
9. Yin, H. & Flynn, A. D. Drugging Membrane Protein Interactions. *Annu. Rev. Biomed. Eng.* **18,** 51–76 (2016).
10. Hardy, D., Bill, R. M., Jawhari, A. & Rothnie, A. J. Overcoming bottlenecks in the membrane protein structural biology pipeline. *Biochem. Soc. Trans.* **44,** 838–844 (2016).
11. Kwan, T. O. C., Axford, D. & Moraes, I. Membrane protein crystallography in the era of modern structural biology. *Biochem. Soc. Trans.* **48,** 2505–2524 (2020).
12. Seddon, A. M., Curnow, P. & Booth, P. J. Membrane proteins, lipids and detergents: not just a soap opera. *Biochim. Biophys. Acta Biomembr.* **1666,** 105–117 (2004).
13. Oliver, R. C. *et al.* Dependence of Micelle Size and Shape on Detergent Alkyl Chain Length and Head Group. *PLoS ONE* **8,** e62488 (2013).
14. Hong, H. & Tamm, L. K. Elastic coupling of integral membrane protein stability to lipid bilayer forces. *PNAS* **101,** 4065–4070 (2004).
15. Jensen, M. Ø. & Mouritsen, O. G. Lipids do influence protein function - the hydrophobic matching hypothesis revisited. *Biochim. Biophys. Acta Biomembr.* **1666,** 205–226 (2004).
16. Andersen, O. S. & Koeppe, R. E. Bilayer Thickness and Membrane Protein Function: An Energetic Perspective. *Annu. Rev. Biophys.* **36,** 107–130 (2007).

17. Epand, R. M., D'Souza, K., Berno, B. & Schlame, M. Membrane curvature modulation of protein activity determined by NMR. *Biochim. Biophys. Acta Biomembr.* **1848,** 220–228 (2015).
18. Marsh, D. Lateral Pressure Profile, Spontaneous Curvature Frustration, and the Incorporation and Conformation of Proteins in Membranes. *Biophys. J.* **93,** 3884–3899 (2007).
19. Cantor, R. S. Lateral Pressures in Cell Membranes: A Mechanism for Modulation of Protein Function. *J. Phys. Chem. B* **101,** 1723–1725 (1997).
20. Cantor, R. S. Lipid composition and the lateral pressure profile in bilayers. *Biophys. J.* **76,** 2625–2639 (1999).
21. Cantor, R. S. The influence of membrane lateral pressures on simple geometric models of protein conformational equilibria. *Chem. Phys. Lipids* **101,** 45–56 (1999).
22. Pliotas, C. *et al.* The role of lipids in mechanosensation. *Nat. Struc. Mol. Biol.* **22,** 991–998 (2015).
23. Kapsalis, C. *et al.* Allosteric activation of an ion channel triggered by modification of mechanosensitive nano-pockets. *Nat. Commun.* **10,** 1–14 (2019).
24. Sanders, C. R. & Prosser, R. S. Bicelles: a model membrane system for all seasons? *Structure* **6,** 1227–1234 (1998).
25. Rigaud, J.-L., Pitard, B. & Levy, D. Reconstitution of membrane proteins into liposomes: application to energy-transducing membrane proteins. *Biochim. Biophys. Acta Bioenerg.* **1231,** 223–246 (1995).
26. Tribet, C., Audebert, R. & Popot, J.-L. Amphipols: Polymers that keep membrane proteins soluble in aqueous solutions. *PNAS* **93,** 15047–15050 (1996).
27. Carlson, M. L. *et al.* The Peptidisc, a simple method for stabilizing membrane proteins in detergent-free solution. *eLife* **7,** e34085 (2018).
28. Denisov, I. G. & Sligar, S. G. Nanodiscs for structural and functional studies of membrane proteins. *Nat. Struct. Mol. Biol.* **23,** 481–486 (2016).
29. Ratkeviciute, G., Cooper, B. F. & Knowles, T. J. Methods for the solubilisation of membrane proteins: the micelle-aneous world of membrane protein solubilisation. *Biochem. Soc. Trans.* **49,** 1763–1777 (2021).
30. Shinzawa-Itoh, K. *et al.* Structures and physiological roles of 13 integral lipids of bovine heart cytochrome *c* oxidase. *EMBO J.* **26,** 1713–1725 (2007).
31. Gupta, K. *et al.* The role of interfacial lipids in stabilizing membrane protein oligomers. *Nature* **541,** 421–424 (2017).
32. Pyle, E. *et al.* Structural Lipids Enable the Formation of Functional Oligomers of the Eukaryotic Purine Symporter UapA. *Cell Chem. Biol.* **25,** 840-848.e4 (2018).

33. Song, W., Yen, H.-Y., Robinson, C. V. & Sansom, M. S. P. State-dependent Lipid Interactions with the A_{2a} Receptor Revealed by MD Simulations Using *In Vivo*-Mimetic Membranes. *Structure* **27,** 392-403.e3 (2019).
34. Sejdiu, B. I. & Tieleman, D. P. Lipid-Protein Interactions Are a Unique Property and Defining Feature of G Protein-Coupled Receptors. *Biophys. J.* **118,** 1887–1900 (2020).
35. Knowles, T. J. *et al.* Membrane Proteins Solubilized Intact in Lipid Containing Nanoparticles Bounded by Styrene Maleic Acid Copolymer. *J. Am. Chem. Soc.* **131,** 7484–7485 (2009).
36. Lee, S. C. *et al.* A method for detergent-free isolation of membrane proteins in their local lipid environment. *Nat. Protoc.* **11,** 1149–1162 (2016).
37. Hesketh, S. J. *et al.* Styrene maleic-acid lipid particles (SMALPs) into detergent or amphipols: An exchange protocol for membrane protein characterisation. *Biochim. Biophys. Acta Biomembr.* **1862,** 183192 (2020).
38. Taylor, T. J. & Vaisman, I. I. Discrimination of thermophilic and mesophilic proteins. *BMC Struct. Biol.* **10,** S5 (2010).
39. Alford, R. F., Fleming, P. J., Fleming, K. G. & Gray, J. J. Protein Structure Prediction and Design in a Biologically Realistic Implicit Membrane. *Biophys. J.* **118,** 2042–2055 (2020).
40. Scott, D. J., Kummer, L., Tremmel, D. & Plückthun, A. Stabilizing membrane proteins through protein engineering. *Curr. Opin. Chem. Biol.* **17,** 427–435 (2013).
41. Heydenreich, F. M., Vuckovic, Z., Matkovic, M. & Veprintsev, D. B. Stabilization of G protein-coupled receptors by point mutations. *Front. Pharmacol.* **6,** (2015).
42. Pace, C. N. & Scholtz, J. M. A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. *Biophysical Journal* **75,** 422–427 (1998).
43. Zhou, Y. & Bowie, J. U. Building a Thermostable Membrane Protein. *J. Biol. Chem.* **275,** 6975–6979 (2000).
44. Schütz, M. *et al.* Directed evolution of G protein-coupled receptors in yeast for higher functional production in eukaryotic expression hosts. *Sci. Rep.* **6,** 21508 (2016).
45. Hattori, M., Hibbs, R. E. & Gouaux, E. A fluorescence-detection size-exclusion chromatography-based thermostability assay to identify membrane protein expression and crystallization conditions. *Structure* **20,** 1293–1299 (2012).
46. Nji, E., Chatzikyriakidou, Y., Landreh, M. & Drew, D. An engineered thermal-shift screen reveals specific lipid preferences of eukaryotic and prokaryotic membrane proteins. *Nat. Commun.* **9,** 1–12 (2018).

47. Toyoshima, C., Nakasako, M., Nomura, H. & Ogawa, H. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature* **405,** 647–655 (2000).
48. Long, S. B., Campbell, E. B. & MacKinnon, R. Crystal Structure of a Mammalian Voltage-Dependent *Shaker* Family K^+ Channel. *Science* **309,** 897–903 (2005).
49. Kurisu, G., Zhang, H., Smith, J. L. & Cramer, W. A. Structure of the Cytochrome b_6f Complex of Oxygenic Photosynthesis: Tuning the Cavity. *Science* **302,** 1009–1014 (2003).
50. Guan, L., Smirnova, I. N., Verner, G., Nagamori, S. & Kaback, H. R. Manipulating phospholipids for crystallization of a membrane transport protein. *PNAS* **103,** 1723–1726 (2006).
51. Cherezov, V. *et al.* High Resolution Crystal Structure of an Engineered Human β_2-Adrenergic G protein-Coupled Receptor. *Science* **318,** 1258–1265 (2007).
52. Nakane, T. *et al.* Single-particle cryo-EM at atomic resolution. *Nature* **587,** 152–156 (2020).
53. Markwick, P. R. L., Malliavin, T. & Nilges, M. Structural Biology by NMR: Structure, Dynamics, and Interactions. *PLoS Comput. Biol.* **4,** e1000168 (2008).
54. Bernhard, R. *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*. (Garland Science, 2019).
55. Bragg, W. H. The reflection of X-rays by crystals. (II.). *Proc. R. Soc. Lond.* **89,** 246–248 (1913).
56. Ishchenko, A., Abola, E. E. & Cherezov, V. in *Protein Crystallography: Methods and Protocols* (eds. Wlodawer, A., Dauter, Z. & Jaskolski, M.) 117–141 (Springer, 2017). doi:10.1007/978-1-4939-7000-1_5
57. Bill, R. M. *et al.* Overcoming barriers to membrane protein structure determination. *Nat. Biotechnol.* **29,** 335–340 (2011).
58. Chun, E. *et al.* Fusion Partner Toolchest for the Stabilization and Crystallization of G Protein-Coupled Receptors. *Structure* **20,** 967–976 (2012).
59. Löw, C. *et al.* Nanobody Mediated Crystallization of an Archeal Mechanosensitive Channel. *PLoS ONE* **8,** e77984 (2013).
60. Kantardjieff, K. A. & Rupp, B. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci.* **12,** 1865–71 (2003).
61. Landau, E. M. & Rosenbusch, J. P. Lipidic cubic phases: A novel concept for the crystallization of membrane proteins. *PNAS* **93,** 14532–14535 (1996).
62. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nat. Protoc.* **4,** 706–731 (2009).

63. Caffrey, M. On the Mechanism of Membrane Protein Crystallization in Lipidic Mesophases. *Cryst. Growth Des.* **8,** 4244–4254 (2008).
64. Misquitta, L. V. *et al.* Membrane protein crystallization in lipidic mesophases with tailored bilayers. *Structure* **12,** 2113–2124 (2004).
65. Rasmussen, S. G. F. *et al.* Crystal structure of the β_2 adrenergic receptor-Gs protein complex. *Nature* **477,** 549–55 (2011).
66. Li, D. & Caffrey, M. Structure and Functional Characterization of Membrane Integral Proteins in the Lipid Cubic Phase. *J. Mol. Biol.* **432,** 5104–5123 (2020).
67. Ma, P. *et al.* The cubicon method for concentrating membrane proteins in the cubic mesophase. *Nat. Protoc.* **12,** 1745–1762 (2017).
68. Gourdon, P. *et al.* HiLiDe - Systematic Approach to Membrane Protein Crystallization in Lipid and Detergent. *Crystal Growth & Design* **11,** 2098–2106 (2011).
69. Lahti, R. Microbial Inorganic Pyrophosphatases. *Microbiological Reviews* **47,** 169–178 (1983).
70. Heinonen, J. K. *Biological Role of Inorganic Pyrophosphate*. (Springer, 2001). doi:10.1007/978-1-4615-1433-6
71. Lohmann, K. Über das Vorkommen und den Umsatz von Pyrophosphat im Muskel. *Naturwissenschaften* **16,** 298–298 (1928).
72. Kay, H. D. The phosphatases of mammalian tissues. *Biochem. J.* **22,** 1446–1448 (1928).
73. Kajander, T., Kellosalo, J. & Goldman, A. Inorganic pyrophosphatases: One substrate, three mechanisms. *FEBS Lett.* **587,** 1863–1869 (2013).
74. Bunick, G., McKenna, P., Scarbrough, F. E., Uberbacher, E. C. & Voet, D. The X-ray structure of yeast inorganic pyrophosphatase at 5.5 Å resolution. *Acta Cryst. B* **34,** 3210–3215 (1978).
75. Cooperman, B. S., Baykov, A. A. & Lahti, R. Evolutionary conservation of the active site of soluble inorganic pyrophosphatase. *Trends Biochem. Sci.* **17,** 262–266 (1992).
76. Josse, J. & Wong, S. C. K. Inorganic Pyrophosphatase of *Escherichia coli*. *Enzymes* **4,** 499–527 (1971).
77. Heikinheimo, P. *et al.* The structural basis for pyrophosphatase catalysis. *Structure* **4,** 1491–1508 (1996).
78. Heikinheimo, P. *et al.* Toward a quantum-mechanical description of metal-assisted phosphoryl transfer in pyrophosphatase. *PNAS* **98,** 3121–3126 (2001).
79. Merckel, M. C. *et al.* Crystal Structure of Streptococcus mutans Pyrophosphatase: A New Fold for an Old Mechanism. *Structure* **9,** 289–297 (2001).

80. Ahn, S. *et al.* The “open” and “closed” structures of the type-C inorganic pyrophosphatases from *Bacillus subtilis* and *Streptococcus gordonii*. *J. Mol. Biol.* **313,** 797–811 (2001).
81. Fabrichniy, I. P. *et al.* A Trimetal Site and Substrate Distortion in a Family II Inorganic Pyrophosphatase. *J. Biol. Chem.* **282,** 1422–1431 (2007).
82. Gonzalez, M. A., Webb, M. R., Welsh, K. M. & Cooperman, B. S. Evidence that catalysis by yeast inorganic pyrophosphatase proceeds by direct phosphoryl transfer to water and not via a phosphoryl enzyme intermediate. *Biochemistry* **23,** 797–801 (1984).
83. Salminen, T. *et al.* Structure and function analysis of *Escherichia coli* inorganic pyrophosphatase: is a hydroxide ion the key to catalysis? *Biochemistry* **34,** 782–791 (1995).
84. Baltscheffsky, H., Von Stedingk, L. V., Heldt, H. W. & Klingenberg, M. Inorganic Pyrophosphate: Formation in Bacterial Photophosphorylation. *Science* **153,** 1120–1122 (1966).
85. Moyle, J., Mitchell, R. & Mitchell, P. Proton-translocating Pyrophosphatase of *Rhodospirillum rubrum*. *FEBS Lett.* **23,** 233–236 (1972).
86. Bucke, C. The distribution and properties of alkaline inorganic pyrophosphatase from higher plants. *Phytochemistry* **9,** 1303–1309 (1970).
87. Karlsson, J. Membrane-bound potassium and magnesium ion-stimulated inorganic pyrophosphatase from roots and cotyledons of sugar beet (*Beta Vulgaris* L.). *Biochim. Biophys. Acta Gen. Subj.* **399,** 356–363 (1975).
88. Scott, D. A. *et al.* Presence of a Plant-like Proton-pumping Pyrophosphatase in Acidocalcisomes of *Trypanosoma cruzi*. *J. Biol. Chem.* **273,** 22151–22158 (1998).
89. Drozdowicz, Y. M. *et al.* A thermostable vacuolar-type membrane pyrophosphatase from the archaeon *Pyrobaculum aerophilum*: implications for the origins of pyrophosphate-energized pumps. *FEBS Lett.* **460,** 505–512 (1999).
90. Malinen, A. M., Belogurov, G. A., Baykov, A. A. & Lahti, R. Na^+-Pyrophosphatase: A Novel Primary Sodium Pump. *Biochemistry* **46,** 8872–8878 (2007).
91. Luoto, H. H., Baykov, A. A., Lahti, R. & Malinen, A. M. Membrane-integral pyrophosphatase subfamily capable of translocating both Na^+ and H^+. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 1255–1260 (2013).
92. Nordbo, E., Luoto, H. H., Baykov, A. A., Lahti, R. & Malinen, A. M. Two independent evolutionary routes to Na^+/H^+ cotransport function in membrane pyrophosphatases. *Biochem. J.* **473,** 3099–3111 (2016).

93. Walker, R. R. & Leigh, R. A. Mg^{2+}-Dependent, cation-stimulated inorganic pyrophosphatase associated with vacuoles isolated from storage roots of red beet (*Beta vulgaris* L.). *Planta* **153,** 150–155 (1981).
94. Ballesteros, J. A. & Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Meth. Neu. Sci.* **25,** 366–428 (1995).
95. Luoto, H. H., Nordbo, E., Malinen, A. M., Baykov, A. A. & Lahti, R. Evolutionarily divergent, Na^+-regulated H^+-transporting membrane-bound pyrophosphatases. *Biochem. J.* **467,** 281–291 (2015).
96. Holmes, A. O. M., Kalli, A. C. & Goldman, A. The Function of Membrane Integral Pyrophosphatases From Whole Organism to Single Molecule. *Front. Mol. Biosci.* **6,** (2019).
97. Yoon, H.-S., Kim, S.-Y. & Kim, I.-S. Stress response of plant H^+-PPase-expressing transgenic *Escherichia coli* and *Saccharomyces cerevisiae*: A potentially useful mechanism for the development of stress-tolerant organisms. *J. Appl. Genet.* **54,** 129–133 (2013).
98. Maeshima, M. TONOPLAST TRANSPORTERS: Organization and Function. *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* **52,** 469–497 (2001).
99. Li, J. *et al.* Arabidopsis H^+-PPase AVP1 Regulates Auxin-Mediated Organ Development. *Science* **310,** 121–125 (2005).
100. Ferjani, A. *et al.* Keep an Eye on PP_i: The Vacuolar-Type H^+-Pyrophosphatase Regulates Postgerminative Development in Arabidopsis. *Plant Cell* **23,** 2895–2908 (2011).
101. Asaoka, M., Segami, S., Ferjani, A. & Maeshima, M. Contribution of PP_i-Hydrolyzing Function of Vacuolar H^+-Pyrophosphatase in Vegetative Growth of Arabidopsis: Evidenced by Expression of Uncoupling Mutated Enzymes. *Front. Plant Sci.* **7,** 415 (2016).
102. Segami, S. *et al.* Vacuolar H^+-Pyrophosphatase and Cytosolic Soluble Pyrophosphatases Cooperatively Regulate Pyrophosphate Levels in Arabidopsis thaliana. *Plant Cell* **30,** 1040–1061 (2018).
103. Fukuda, M. *et al.* Lack of Vacuolar H^+-Pyrophosphatase and Cytosolic Pyrophosphatases Causes Fatal Developmental Defects in *Arabidopsis thaliana*. *Front Plant Sci.* **11,** 655 (2020).
104. Gaxiola, R. A. *et al.* Drought- and salt-tolerant plants result from overexpression of the AVP1 H^+-pump. *PNAS* **98,** 11444–11449 (2001).
105. Park, S. *et al.* Up-regulation of a H^+-pyrophosphatase (H^+-PPase) as a strategy to engineer drought-resistant crop plants. *PNAS* **102,** 18830–18835 (2005).

106. Lv, S. *et al.* Overexpression of an H^+-PPase gene from *Thellungiella halophila* in cotton enhances salt tolerance and improves growth and photosynthetic performance. *Plant Cell Physiol.* **49,** 1150–1164 (2008).
107. Lv, S.-L. *et al.* Overexpression of *Thellungiella halophila* H(+)-PPase (TsVP) in cotton enhances drought stress resistance of plants. *Planta* **229,** 899–910 (2009).
108. Bao, A.-K. *et al.* Overexpression of the Arabidopsis H^+-PPase enhanced resistance to salt and drought stress in transgenic alfalfa (*Medicago sativa L.*). *Plant Sci.* **176,** 232–240 (2009).
109. Pasapula, V. *et al.* 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,** 88–99 (2011).
110. Asif, M. A. *et al.* Enhanced Expression of *AtNHX1*, in Transgenic Groundnut (*Arachis hypogaea* L.) Improves Salt and Drought Tolerence. *Mol. Biotechnol.* **49,** 250–256 (2011).
111. Kim, Y. S. *et al.* Overexpression of the Arabidopsis vacuolar H^+-pyrophosphatase AVP1 gene in rice plants improves grain yield under paddy field conditions. *J. Agric. Sci.* **152,** 941–953 (2014).
112. Esmaeili, N. *et al.* Co-overexpression of *AVP1* and *OsSIZ1* in Arabidopsis substantially enhances plant tolerance to drought, salt, and heat stresses. *Sci. Rep.* **9,** 1–15 (2019).
113. Schulze, W. X., Schneider, T., Starck, S., Martinoia, E. & Trentmann, O. Cold acclimation induces changes in Arabidopsis tonoplast protein abundance and activity and alters phosphorylation of tonoplast monosaccharide transporters. *Plant J.* **69,** 529–541 (2012).
114. Docampo, R. & Moreno, S. N. J. Acidocalcisomes. *Cell Calcium* **50,** 113–119 (2011).
115. Lander, N., Cordeiro, C., Huang, G. & Docampo, R. Polyphosphate and acidocalcisomes. *Biochem. Soc. Trans.* **44,** 1–6 (2016).
116. Blum, J. J., Ingram, P. & LeFurgey, A. Compartmental responses to acute osmotic stress in Leishmania major result in rapid loss of Na+ and Cl-. *Comparative biochemistry and physiology* **128,** 385–394 (2001).
117. Ruiz, F. A., Rodrigues, C. O. & Docampo, R. Rapid Changes in Polyphosphate Content within Acidocalcisomes in Response to Cell Growth, Differentiation, and Environmental Stress in *Trypanosoma cruzi*. *J. Biol. Chem.* **276,** 26114–26121 (2001).
118. Lemercier, G. *et al.* 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,** 37369–37376 (2002).

119. Nakanishi, Y., Saijo, T., Wada, Y. & Maeshima, M. Mutagenic Analysis of Functional Residues in Putative Substrate-binding Site and Acidic Domains of Vacuolar H^+-Pyrophosphatase. *J. Biol. Chem.* **276,** 7654–7660 (2001).
120. Schultz, A. & Baltscheffsky, M. Properties of mutated Rhodospirillum rubrum H^+-pyrophosphatase expressed in *Escherichia coli*. *Biochim. Biophys. Acta Bioenerg.* **1607,** 141–151 (2003).
121. Hirono, M., Nakanishi, Y. & Maeshima, M. Essential amino acid residues in the central transmembrane domains and loops for energy coupling of *Streptomyces coelicolor* A3(2) H^+-pyrophosphatase. *Biochim. Biophys. Acta Bioenerg.* **1767,** 930–939 (2007).
122. Pan, Y.-J. *et al.* The transmembrane domain 6 of vacuolar H^+-pyrophosphatase mediates protein targeting and proton transport. *Biochim. Biophys. Acta Bioenerg.* **1807,** 59–67 (2011).
123. Zhen, R.-G., Kim, E. J. & Rea, P. A. Acidic Residues Necessary for Pyrophosphate-energized Pumping and Inhibition of the Vacuolar H^+-pyrophosphatase by *N,N'*-Dicyclohexylcarbodiimide. *J. Biol. Chem.* **272,** 22340–22348 (1997).
124. Mimura, H., Nakanishi, Y., Hirono, M. & Maeshima, M. Membrane Topology of the H^+-pyrophosphatase of *Streptomyces coelicolor* Determined by Cysteine-scanning Mutagenesis. *J. Biol. Chem.* **279,** 35106–35112 (2004).
125. Hirono, M. & Maeshima, M. Functional Enhancement by Single-residue Substitution of *Streptomyces coelicolor* A3(2) H^+-translocating Pyrophosphatase. *J. Biochem.* **146,** 617–621 (2009).
126. Luoto, H. H., Belogurov, G. A., Baykov, A. A., Lahti, R. & Malinen, A. M. Na^+-translocating Membrane Pyrophosphatases Are Widespread in the Microbial World and Evolutionarily Precede H^+-translocating Pyrophosphatases. *The Journal of Biological Chemistry* **286,** 21633–21642 (2011).
127. Kellosalo, J., Kajander, T., Kogan, K., Pokharel, K. & Goldman, A. The Structure and Catalytic Cycle of a Sodium-Pumping Pyrophosphatase. *Science* **337,** 473–476 (2012).
128. Lin, S.-M. *et al.* Crystal structure of a membrane-embedded H^+-translocating pyrophosphatase. *Nature* **484,** 399–404 (2012).
129. Li, K.-M. *et al.* Membrane pyrophosphatases from *Thermotoga maritima* and *Vigna radiata* suggest a conserved coupling mechanism. *Nat. Commun.* **7,** 1–11 (2016).
130. Tsai, J.-Y. *et al.* Roles of the Hydrophobic Gate and Exit Channel in Vigna radiata Pyrophosphatase Ion Translocation. *J. Mol. Biol.* **431,** 1619–1632 (2019).
131. Vidilaseris, K. *et al.* Asymmetry in catalysis by *Thermotoga maritima* membrane-bound pyrophosphatase demonstrated by a nonphosphorus allosteric inhibitor. *Sci. Adv.* **5,** eaav7574 (2019).

132. Asaoka, M., Segami, S. & Maeshima, M. Identification of the critical residues for the function of vacuolar H^+-pyrophosphatase by mutational analysis based on the 3D structure. *J. Biochem.* **156,** 333–344 (2014).
133. Maeshima, M. Vacuolar H^+-pyrophosphatase. *Biochim. Biophys. Acta Biomembr.* **1465,** 37–51 (2000).
134. Malinen, A. M., Baykov, A. A. & Lahti, R. Mutual Effects of Cationic Ligands and Substrate on Activity of the Na^+-Transporting Pyrophosphatase of *Methanosarcina mazei*. *Biochemistry* **47,** 13447–13454 (2008).
135. Rea, P. A. *et al.* Regulation of Vacuolar H^+-Pyrophosphatase by Free Calcium. *Plant Physiol.* **100,** 1706–1715 (1992).
136. Artukka, E., Luoto, H. H., Baykov, A. A., Lahti, R. & Malinen, A. M. Role of the potassium/lysine cationic center in catalysis and functional asymmetry in membrane-bound pyrophosphatases. *Biochem. J.* **475,** 1141–1158 (2018).
137. Anashkin, V. A., Malinen, A. M., Bogachev, A. & Baykov, A. A. Catalytic Asymmetry in Homodimeric H^+ -Pumping Membrane Pyrophosphatase Demonstrated by Non-Hydrolyzable Pyrophosphate Analogs. *Int. J. Mol. Sci.* **22,** 9820 (2021).
138. Achbergerová, L. & Nahálka, J. Polyphosphate - an ancient energy source and active metabolic regulator. *Microbial Cell Factories* **10,** 1–14 (2011).
139. Mulkidjanian, A. Y., Galperin, M. Y., Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Evolutionary primacy of sodium bioenergetics. *Biology Direct* **3,** 1–19 (2008).
140. Mulkidjanian, A. Y., Dibrov, P. & Galperin, M. Y. The past and present of the sodium energetics: May the sodium-motive force be with you. *Biochimica et Biophysica Acta - Bioenergetics* **1777,** 985–992 (2008).
141. Belogurov, G. A. & Lahti, R. A Lysine Substitute for K^+. *J. Biol. Chem.* **277,** 49651–49654 (2002).
142. Shah, N. R. *et al.* Insights into the mechanism of membrane pyrophosphatases by combining experiment and computer simulation. *Struct. Dyn.* **4,** 1–12 (2017).
143. Lee, C.-H. *et al.* Identification of Essential Lysines Involved in Substrate Binding of Vacuolar H+-Pyrophosphatase. *J. Biol. Chem.* **286,** 11970–11976 (2011).
144. Baykov, A. A. Energy Coupling in Cation-Pumping Pyrophosphatase - Back to Mitchell. *Front. Plant Sci.* **11,** (2020).
145. Baykov, A. A., Malinen, A. M., Luoto, H. H. & Lahti, R. Pyrophosphate-Fueled Na^+ and H^+ Transport in Prokaryotes. *Microbiol. Mol. Biol. Rev.* **77,** 267–276 (2013).
146. Polgár, L. The mechanism of action of aspartic proteases involves 'push-pull' catalysis. *FEBS Lett.* **219,** 1–4 (1987).
147. Mitchell, P. A chemiosmotic molecular mechanism for proton-translocating adenosine triphosphatases. *FEBS Letters* **43,** 189–194 (1974).

148. Boyer, P. D. The ATP synthase - a splendid molecular machine. *Annu. Rev. Biochem.* **66,** 717–749 (1997).
149. Wu, J. J., Ma, J. T. & Pan, R. L. Functional size analysis of pyrophosphatase from *Rhodospirillum rubrum* determined by radiation inactivation. *FEBS Lett.* **283,** 57–60 (1991).
150. Sarafian, V., Potier, M. & Poole, R. J. Radiation-inactivation analysis of vacuolar H^+-ATPase and H^+-pyrophosphatase from *Beta vulgaris* L. Functional sizes for substrate hydrolysis and for H^+ transport. *Biochemical Journal* **283,** 493–497 (1992).
151. Tzeng, C. M. *et al.* Subunit structure of vacuolar proton-pyrophosphatase as determined by radiation inactivation. *Biochem. J.* **316,** 143–147 (1996).
152. Pérez-Castiñeira, J. R. & Serrano, A. The H^+-Translocating Inorganic Pyrophosphatase From Arabidopsis thaliana Is More Sensitive to Sodium Than Its Na^+-Translocating Counterpart From *Methanosarcina mazei*. *Front Plant Sci.* **11,** 1240 (2020).
153. Yang, S. J., Jiang, S. S., Van, R. C., Hsiao, Y. Y. & Pan, R.-L. A lysine residue involved in the inhibition of vacuolar H^+-pyrophosphatase by fluorescein 5′-isothiocyanate. *Biochim. Biophys. Acta Bioenerg.* **1460,** 375–383 (2000).
154. García-Contreras, R., Celis, H. & Romero, I. Importance of *Rhodospirillum rubrum* H^+-Pyrophosphatase under Low-Energy Conditions. *J. Bacteriol.* **186,** 6651–6655 (2004).
155. Battisti, D. S. & Naylor, R. L. Historical Warnings of Future Food Insecurity with Unprecedented Seasonal Heat. *Science* **323,** 240–244 (2009).
156. Mazdiyasni, O. & AghaKouchak, A. Substantial increase in concurrent droughts and heatwaves in the United States. *PNAS* **112,** 11484–11489 (2015).
157. Lesk, C., Rowhani, P. & Ramankutty, N. Influence of extreme weather disasters on global crop production. *Nature* **529,** 84–87 (2016).
158. Zhao, C. *et al.* Temperature increase reduces global yields of major crops in four independent estimates. *PNAS* **114,** 9326–9331 (2017).
159. Ray, D. K. *et al.* Climate change has likely already affected global food production. *PLoS ONE* **14,** e0217148 (2019).
160. Ryan, S. J., Carlson, C. J., Mordecai, E. A. & Johnson, L. R. Global expansion and redistribution of Aedes-borne virus transmission risk with climate change. *PLoS Negl. Trop. Dis.* **13,** (2019).
161. Hertig, E. Distribution of *Anopheles* vectors and potential malaria transmission stability in Europe and the Mediterranean area under future climate change. *Parasites Vectors* **12,** (2019).

162. Zhang, M. *et al.* Uncovering the essential genes of the human malaria parasite *Plasmodium falciparum* by saturation mutagenesis. *Science* **360,** (2018).
163. Lemercier, G. *et al.* 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,** 37369–37376 (2002).
164. Liu, J. *et al.* A Vacuolar-H^+-Pyrophosphatase (TgVP1) is Required for Microneme Secretion, Host Cell Invasion, and Extracellular Survival of *Toxoplasma gondii*. *Mol. Microbiol.* **93,** 698–712 (2014).
165. Wexler, H. M. Bacteroides: the Good, the Bad, and the Nitty-Gritty. *Clin. Microbiol. Rev.* **20,** 593–621 (2007).
166. United Nations Interagency Coordination Group on Antimicrobial Resistance. *No Time to Wait: Securing the Future from Drug-Resistant Infections*. 28 (United Nations Interagency Coordination Group on Antimicrobial Resistance, 2019). at <https://repository.gheli.harvard.edu/repository/12976/>
167. Johansson, N. G. *et al.* Exploration of Pyrazolo[1,5-a]pyrimidines as Membrane-Bound Pyrophosphatase Inhibitors. *ChemMedChem* 1–9 (2021). doi:10.1002/cmdc.202100392
168. Johansson, N. G. *et al.* Discovery of Membrane-Bound Pyrophosphatase Inhibitors Derived from an Isoxazole Fragment. *ACS Med. Chem. Lett.* (2020). doi:10.1021/acsmedchemlett.9b00537
169. Vidilaseris, K., Kellosalo, J. & Goldman, A. A high-throughput method for orthophosphate determination of thermostable membrane-bound pyrophosphatase activity. *Anal. Methods* **10,** 646–651 (2018).
170. Nagul, E. A., McKelvie, I. D., Worsfold, P. & Kolev, S. D. The molybdenum blue reaction for the determination of orthophosphate revisited: Opening the black box. *Analytica Chimica Acta* **890,** 60–82 (2015).
171. Baykov, A. A., Bakuleva, N. P. & Rea, P. A. Steady-state kinetics of substrate hydrolysis by vacuolar H^+-pyrophosphatase. *FEBS J.* **217,** 755–762 (1993).
172. Cecchetti, C. *et al.* A novel high-throughput screen for identifying lipids that stabilise membrane proteins in detergent based solution. *PLOS ONE* **16,** e0254118 (2021).
173. Harborne, S. P. D. *et al.* IMPROvER: the Integral Membrane Protein Stability Selector. *Sci. Rep.* **10,** 15165 (2020).
174. Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinformatics* **54,** 5.6.1-5.6.37 (2016).
175. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5,** 725–738 (2010).

176. Drew, D. *et al.* GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae*. *Nat. Protoc.* **3,** 784–798 (2008).
177. López-Marqués, R. L. *et al.* 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 Biomembr.* **1716,** 69–76 (2005).
178. Wilkinson, C. Understanding the Catalytic Cycle of Membrane Pyrophosphatases Through Structural and Functional Studies. (2017). at <http://etheses.whiterose.ac.uk/19131/>
179. Stohrer, C. *et al.* Homogeneous batch micro-crystallization of proteins from ammonium sulfate. *Acta Cryst. D* **77,** 194–204 (2021).
180. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Cryst. D* **66,** 133–144 (2010).
181. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40,** 658–674 (2007).
182. Tickle, I. J. *et al.* The STARANISO Server: Anisotropy of the Diffraction Limit andBayesian Estimation of Structure Amplitudes. *The STARANISO Server* (2018).
183. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Cryst. D* **75,** 861–877 (2019).
184. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *Coot*. *Acta Cryst. D* **66,** 486–501 (2010).
185. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
186. Schrödinger, LLC. *The PyMOL Molecular Graphics System, Version 2.0*. at <https://www.pymol.org/>
187. Ho, B. K. & Gruswitz, F. HOLLOW: Generating Accurate Representations of Channel and Interior Surfaces in Molecular Structures. *BMC Struct. Biol.* **8,** 49 (2008).
188. Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A. & Caves, L. S. D. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22,** 2695–2696 (2006).
189. McDonald, I. K. & Thornton, J. M. Satisfying Hydrogen Bonding Potential in Proteins. *J. Mol. Biol.* **238,** 777–793 (1994).
190. Dahl, A. C. E., Chavent, M. & Sansom, M. S. P. Bendix: intuitive helix geometry analysis and abstraction. *Bioinformatics* **28,** 2193–2194 (2012).

191. Kumar, P. & Bansal, M. HELANAL-Plus: a web server for analysis of helix geometry in protein structures. *J. Biomol. Struct. Dyn.* **30,** 773–783 (2012).
192. Mehrabi, P. *et al.* The HARE chip for efficient time-resolved serial synchrotron crystallography. *J. Synchrotron Radiat.* **27,** 360–370 (2020).
193. Gevorkov, Y. *et al.* XGANDALF - extended gradient descent algorithm for lattice finding. *Acta Cryst. A* **75,** 694–704 (2019).
194. White, T. A. *et al.* CrystFEL: a software suite for snapshot serial crystallography. *J. Appl. Cryst.* **45,** 335–341 (2012).
195. Winter, G. *et al.* DIALS: implementation and evaluation of a new integration package. *Acta Cryst. D* **74,** 85–97 (2018).
196. Foadi, J. *et al.* Clustering procedures for the optimal selection of data sets from multiple crystals in macromolecular crystallography. *Acta Cryst. D* **69,** 1617–1632 (2013).
197. Lundbæk, J. A., Collingwood, S. A., Ingólfsson, H. I., Kapoor, R. & Andersen, O. S. Lipid bilayer regulation of membrane protein function: gramicidin channels as molecular force probes. *J. Royal Soc. Interface* **7,** 373–395 (2010).
198. Laganowsky, A. *et al.* Membrane proteins bind lipids selectively to modulate their structure and function. *Nature* **510,** 172–175 (2014).
199. Potapov, V., Cohen, M. & Schreiber, G. Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. *Protein Eng. Des. Sel.* **22,** 553–560 (2009).
200. Magnani, F., Shibata, Y., Serrano-Vega, M. J. & Tate, C. G. Co-evolving stability and conformational homogeneity of the human adenosine A_{2a} receptor. *PNAS* **105,** 10744–10749 (2008).
201. Warne, T., Serrano-Vega, M. J., Tate, C. G. & Schertler, G. F. X. Development and crystallization of a minimal thermostabilised G protein-coupled receptor. *Protein Expr. Purif.* **65,** 204–213 (2009).
202. Shibata, Y. *et al.* Optimising the combination of thermostabilising mutations in the neurotensin receptor for structure determination. *Biochim. Biophys. Acta Biomembr.* **1828,** 1293–1301 (2013).
203. Yasuda, S. *et al.* Identification of Thermostabilizing Mutations for Membrane Proteins: Rapid Method Based on Statistical Thermodynamics. *J. Phys. Chem. B* **120,** 3833–3843 (2016).
204. Kellosalo, J., Kajander, T., Palmgren, M. G., Lopéz-Marqués, R. L. & Goldman, A. Heterologous expression and purification of membrane-bound pyrophosphatases. *Protein Expr. Purif.* **79,** 25–34 (2011).
205. Ashok, Y., Nanekar, R. & Jaakola, V.-P. Defining thermostability of membrane proteins by western blotting. *Protein Eng. Des. Sel.* **28,** 539–542 (2015).

206. Hénault, C. M. *et al.* A lipid site shapes the agonist response of a pentameric ligand-gated ion channel. *Nat. Chem. Biol.* **15,** 1156–1164 (2019).
207. McGoldrick, L. L. *et al.* Structure of the thermo-sensitive TRP channel TRP1 from the alga *Chlamydomonas reinhardtii*. *Nat. Commun.* **10,** 4180 (2019).
208. Vitrac, H., Mallampalli, V. K. P. S., Bogdanov, M. & Dowhan, W. The lipid-dependent structure and function of LacY can be recapitulated and analyzed in phospholipid-containing detergent micelles. *Sci. Rep.* **9,** 11338 (2019).
209. Broecker, J., Eger, B. T. & Ernst, O. P. Crystallogenesis of Membrane Proteins Mediated by Polymer-Bounded Lipid Nanodiscs. *Structure* **25,** 384–392 (2017).
210. Srivastava, S. R., Zadafiya, P. & Mahalakshmi, R. Hydrophobic Mismatch Modulates Stability and Plasticity of Human Mitochondrial VDAC2. *Biophys. J.* **115,** 2386–2394 (2018).
211. Baturin, S., Galka, J. J., Piyadasa, H., Gajjeraman, S. & O'Neil, J. D. The effects of a protein osmolyte on the stability of the integral membrane protein glycerol facilitator. *Biochem. Cell Biol.* **92,** 564–575 (2014).
212. Valiyaveetil, F. I., Zhou, Y. & MacKinnon, R. Lipids in the Structure, Folding, and Function of the KcsA K^+ Channel. *Biochemistry* **41,** 10771–10777 (2002).
213. Koshy, C. *et al.* Structural evidence for functional lipid interactions in the betaine transporter BetP. *EMBO J.* **32,** 3096–3105 (2013).
214. Pyle, E. *et al.* Protein–Lipid Interactions Stabilize the Oligomeric State of BOR1p from *Saccharomyces cerevisiae*. *Anal. Chem.* **91,** 13071–13079 (2019).
215. Sharom, F. J. Complex Interplay between the P-Glycoprotein Multidrug Efflux Pump and the Membrane: Its Role in Modulating Protein Function. *Front. Oncol.* **4,** (2014).
216. Laursen, L. *et al.* Cholesterol binding to a conserved site modulates the conformation, pharmacology, and transport kinetics of the human serotonin transporter. *J. Biol. Chem.* **293,** 3510–3523 (2018).
217. Jungnickel, K. E. J., Parker, J. L. & Newstead, S. Structural basis for amino acid transport by the CAT family of SLC7 transporters. *Nat. Commun.* **9,** 550 (2018).
218. Palsdottir, H. & Hunte, C. Lipids in membrane protein structures. *Biochim. Biophys. Acta Biomembr.* **1666,** 2–18 (2004).
219. van Dalen, A., Hegger, S., Killian, J. A. & Kruijff, B. de. Influence of lipids on membrane assembly and stability of the potassium channel KcsA. *FEBS Lett.* **525,** 33–38 (2002).
220. Zubcevic, L. *et al.* Cryo-electron microscopy structure of the TRPV2 ion channel. *Nat. Struct. Mol. Biol.* **23,** 180–186 (2016).
221. Deng, Z. *et al.* Cryo-EM and X-ray structures of TRPV4 reveal insight into ion permeation and gating mechanisms. *Nat. Struct. Mol. Biol.* **25,** 252–260 (2018).

222. Jin, P. *et al.* Electron cryo-microscopy structure of the mechanotransduction channel NOMPC. *Nature* **547,** 118–122 (2017).
223. Igonet, S. *et al.* Enabling STD-NMR fragment screening using stabilized native GPCR: A case study of adenosine receptor. *Sci. Rep.* **8,** 8142 (2018).
224. Kotov, V. *et al.* High-throughput stability screening for detergent-solubilized membrane proteins. *Sci. Rep.* **9,** 10379 (2019).
225. Schrecke, S. *et al.* Selective regulation of human TRAAK channels by biologically active phospholipids. *Nat. Chem. Biol.* **17,** 89–95 (2021).
226. Jaakola, V.-P. *et al.* The 2.6 Angstrom Crystal Structure of a Human A_{2A} Adenosine Receptor Bound to an Antagonist. *Science* **322,** 1211–1217 (2008).
227. Magnani, F. *et al.* Thermostabilisation of membrane proteins for structural studies. *Nat. Protoc.* **11,** 1554–1571 (2016).
228. Sauer, D. B., Karpowich, N. K., Song, J. M. & Wang, D.-N. Rapid Bioinformatic Identification of Thermostabilizing Mutations. *Biophys. J.* **109,** 1420–1428 (2015).
229. Pires, D. E. V., Rodrigues, C. H. M. & Ascher, D. B. mCSM-membrane: predicting the effects of mutations on transmembrane proteins. *Nucleic Acids Research* **48,** W147–W153 (2020).
230. Bhattacharya, S., Lee, S., Grisshammer, R., Tate, C. G. & Vaidehi, N. Rapid Computational Prediction of Thermostabilizing Mutations for G Protein-Coupled Receptors. *J. Chem. Theory Comput.* **10,** 5149–5160 (2014).
231. Popov, P. *et al.* Computational design of thermostabilizing point mutations for G protein-coupled receptors. *eLife* **7,** (2018).
232. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596,** 583–589 (2021).
233. Pucci, F., Bourgeas, R. & Rooman, M. High-quality Thermodynamic Data on the Stability Changes of Proteins Upon Single-site Mutations. *J. Phys. Chem. Ref. Data* **45,** 023104 (2016).
234. Serrano-Vega, M. J., Magnani, F., Shibata, Y. & Tate, C. G. Conformational thermostabilization of the <i>β</>1-adrenergic receptor in a detergent-resistant form. *PNAS* **105,** 877–882 (2008).
235. Muyldermans, S. Nanobodies: Natural Single-Domain Antibodies. *Annu. Rev. Biochem.* **82,** 775–797 (2013).
236. Malawski, G. A. *et al.* Identifying protein construct variants with increased crystallization propensity—A case study. *Protein Sci.* **15,** 2718–2728 (2006).
237. Vedadi, M. *et al.* Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *PNAS* **103,** 15835–15840 (2006).

238. Guan, L., Mirza, O., Verner, G., Iwata, S. & Kaback, H. R. Structural determination of wild-type lactose permease. *PNAS* **104,** 15294–15298 (2007).
239. Kean, J., Bortolato, A., Hollenstein, K., Marshall, F. H. & Jazayeri, A. Conformational thermostabilisation of corticotropin releasing factor receptor 1. *Sci. Rep.* **5,** 11954 (2015).
240. Liu, W. *et al.* Structural Basis for Allosteric Regulation of GPCRs by Sodium Ions. *Science* **337,** 232–236 (2012).
241. Weiß, H. M. & Grisshammer, R. Purification and characterization of the human adenosine A_{2a} receptor functionally expressed in *Escherichia coli*. *Eur. J. Biochem.* **269,** 82–92 (2002).
242. Bruzzese, A., Dalton, J. A. R. & Giraldo, J. Insights into adenosine A_{2A} receptor activation through cooperative modulation of agonist and allosteric lipid interactions. *PLoS Comput. Biol.* **16,** (2020).
243. Zoe Cournia, †, G. Matthias Ullmann, † & Jeremy C. Smith*, †. Differential Effects of Cholesterol, Ergosterol and Lanosterol on a Dipalmitoyl Phosphatidylcholine Membrane: A Molecular Dynamics Simulation Study. *J. Phys. Chem. B* **111,** 1786–1801 (2007).
244. Leung, J., Cameron, A. D., Diallinas, G. & Byrne, B. Stabilizing the heterologously expressed uric acid-xanthine transporter UapA from the lower eukaryote *Aspergillus nidulans*. *Mol. Membr. Biol.* **30,** 32–42 (2013).
245. Tate, C. G. in *Heterologous Expression of Membrane Proteins: Methods and Protocols* (ed. Mus-Veteau, I.) **601,** 187–203 (2010).
246. Heerklotz, H. Interactions of surfactants with lipid membranes. *Q. Rev. Biophys.* **41,** 205–264 (2008).
247. Chae, P. S. *et al.* Glucose-Neopentyl Glycol (GNG) Amphiphiles for Membrane Protein Solubilization, Stabilization and Crystallization. *Chem. Commun.* **49,** 2287–2289 (2013).
248. Zhang, Y. *et al.* Visualization of the mechanosensitive ion channel MscS under membrane tension. *Nature* **590,** 509–514 (2021).
249. Kellosalo, J. Structural studies of membrane –bound pyrophosphatases. (2013). at <https://helda.helsinki.fi/bitstream/handle/10138/40788/kellosalo_dissertation.pdf?isAllowed=y&sequence=1>
250. Lovell, S. C., Word, J. M., Richardson, J. S. & Richardson, D. C. The penultimate rotamer library. *Proteins* **40,** 389–408 (2000).
251. Strop, P. & Brunger, A. T. Refractive index-based determination of detergent concentration and its application to the study of membrane proteins. *Protein Sci.* **14,** 2207–2211 (2005).

252. Urbani, A. & Warne, T. A colorimetric determination for glycosidic and bile salt-based detergents: applications in membrane protein research. *Anal. Biochem.* **336,** 117–124 (2005).
253. Maslennikov, I. *et al.* NMR spectroscopic and analytical ultracentrifuge analysis of membrane protein detergent complexes. *BMC Struct. Biol.* **7,** 74 (2007).
254. Chae, P. S. *et al.* Maltose-neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins. *Nat. Methods* **7,** 1003–1008 (2010).
255. Stetsenko, A. & Guskov, A. An Overview of the Top Ten Detergents Used for Membrane Protein Crystallization. *Crystals* **7,** 197 (2017).
256. Huang, C. Y. *et al. In meso in situ* serial X-ray crystallography of soluble and membrane proteins. *Acta Cryst. D* **71,** 1238–56 (2015).
257. Bandura, A. V. & Lvov, S. N. The Ionization Constant of Water over Wide Ranges of Temperature and Density. *J. Phys. Chem. Ref. Data* **35,** 15–30 (2005).
258. Völkl, P. *et al.* Pyrobaculum aerophilum sp. nov., a novel nitrate-reducing hyperthermophilic archaeum. *Appl. Environ. Microbiol.* **59,** 2918–2926 (1993).
259. Luoto, H. H., Nordbo, E., Baykov, A. A., Lahti, R. & Malinen, A. M. Membrane Na^+-pyrophosphatases Can Transport Protons at Low Sodium Concentrations. *J. Biol. Chem.* **288,** 35489–35499 (2013).
260. Luoto, H. H., Belogurov, G. A., Baykov, A. A., Lahti, R. & Malinen, A. M. Na^+-translocating Membrane Pyrophosphatases Are Widespread in the Microbial World and Evolutionarily Precede H^+-translocating Pyrophosphatases. *J. Biol. Chem.* **286,** 21633–21642 (2011).
261. Tsai, J.-Y., Kellosalo, J., Sun, Y.-J. & Goldman, A. Proton/sodium pumping pyrophosphatases: the last of the primary ion pumps. *Curr. Opin. Struct. Biol.* **27,** 38–47 (2014).
262. Wisedchaisri, G. *et al.* Resting-State Structure and Gating Mechanism of a Voltage-Gated Sodium Channel. *Cell* **178,** 993-1003.e12 (2019).
263. Infield, D. T. *et al.* Main-chain mutagenesis reveals intrahelical coupling in an ion channel voltage-sensor. *Nat. Commun.* **9,** 5055 (2018).
264. Schwaiger, C. S., Bjelkmar, P., Hess, B. & Lindahl, E. 310-Helix Conformation Facilitates the Transition of a Voltage Sensor S4 Segment toward the Down State. *Biophys. J.* **100,** 1446–1454 (2011).
265. Baykov, A. A., Anashkin, V. A. & Malinen, A. M. Good-Practice Non-Radioactive Assays of Inorganic Pyrophosphatase Activities. *Molecules* **26,** 2356 (2021).
266. Abbott, G. A. THE RATE OF HYDRATION OF PYROPHOSPHORIC ACID IN AQUEOUS SOLUTION. *J. Am. Chem. Soc.* **31,** 763–770 (1909).

267. Breibeck, J. & Rompel, A. Successful amphiphiles as the key to crystallization of membrane proteins: Bridging theory and practice. *Biochim. Biophys. Acta Gen. Subj.* **1863,** 437–455 (2019).
268. Newport, T. D., Sansom, M. S. P. & Stansfeld, P. J. The MemProtMD database: a resource for membrane-embedded protein structures and their lipid interactions. *Nucleic Acids Res.* **47,** D390–D397 (2019).
269. Harborne, S. P. D. *et al.* Revolutionising the design and analysis of protein engineering experiments using fractional factorial design. *bioRxiv* 298273 (2018). doi:10.1101/298273
270. Levantino, M., Yorke, B. A., Monteiro, D. CF., Cammarata, M. & Pearson, A. R. Using synchrotrons and XFELs for time-resolved X-ray crystallography and solution scattering experiments on biomolecules. *Curr. Opin. Struct. Biol.* **35,** 41–48 (2015).
271. Schmidt, M. Mix and Inject: Reaction Initiation by Diffusion for Time-Resolved Macromolecular Crystallography. *Adv. Condens. Matter Phys.* **2013,** e167276 (2013).
272. Poddar, H. *et al.* A guide to time-resolved structural analysis of light-activated proteins. *FEBS J.* **n/a,**
273. Kubelka, J. Time-resolved methods in biophysics. 9. Laser temperature-jump methods for investigating biomolecular dynamics. *Photochem. Photobiol. Sci.* **8,** 499–512 (2009).
274. Wang, J. *et al.* Time-resolved protein activation by proximal decaging in living systems. *Nature* **569,** 509–513 (2019).
275. Pearson, A. R. & Mehrabi, P. Serial synchrotron crystallography for time-resolved structural biology. *Current Opinion in Structural Biology* **65,** 168–174 (2020).
276. Martin-Garcia, J. M. *et al.* High-viscosity injector-based pink-beam serial crystallography of microcrystals at a synchrotron radiation source. *IUCrJ* **6,** 412–425 (2019).
277. Pandey, S. *et al.* Time-resolved serial femtosecond crystallography at the European XFEL. *Nat. Methods* **17,** 73–78 (2020).
278. Nogly, P. *et al.* Lipidic cubic phase injector is a viable crystal delivery system for time-resolved serial crystallography. *Nat. Commun.* **7,** 12314 (2016).
279. Monteiro, D. C. F. *et al.* 3D-MiXD: 3D-printed X-ray-compatible microfluidic devices for rapid, low-consumption serial synchrotron crystallography data collection in flow. *IUCrJ* **7,** 207–219 (2020).
280. Mehrabi, P. *et al.* Liquid application method for time-resolved analyses by serial synchrotron crystallography. *Nat. Methods* **16,** (2019).

281. Aller, P., Geng, T., Evans, G. & Foadi, J. in *The Next Generation in Membrane Protein Structure Determination* (ed. Moraes, I.) 119–135 (Springer International Publishing, 2016). doi:10.1007/978-3-319-35072-1_9
282. Latorraca, N. R., Venkatakrishnan, A. J. & Dror, R. O. GPCR Dynamics: Structures in Motion. *Chem. Rev.* **117,** 139–155 (2017).
283. McPherson, A. & Cudney, B. Optimization of crystallization conditions for biological macromolecules. *Acta Cryst. F* **70,** 1445–1467 (2014).
284. Chim, N., Meza, R. A., Trinh, A. M., Yang, K. & Chaput, J. C. Following replicative DNA synthesis by time-resolved X-ray crystallography. *Nat. Commun.* **12,** 2641 (2021).
285. Han, S., Kim, C. & Kwon, D. Thermal/oxidative degradation and stabilization of polyethylene glycol. *Polymer* **38,** 317–323 (1997).
286. Zimmerman, S. B. & Pheiffer, B. H. Macromolecular crowding allows blunt-end ligation by DNA ligases from rat liver or Escherichia coli. *PNAS* **80,** 5852–5856 (1983).
287. Zimmerman, S. B. & Harrison, B. Macromolecular crowding increases binding of DNA polymerase to DNA: an adaptive effect. *PNAS* **84,** 1871–1875 (1987).
288. McMullan, G., Faruqi, A. R. & Henderson, R. in *Methods in Enzymology* (ed. Crowther, R. A.) **579,** 1–17 (Academic Press, 2016).
289. Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife* **5,** e18722 (2016).
290. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14,** 290–296 (2017).
291. Klebl, D. P. *et al.* Need for Speed: Examining Protein Behavior during CryoEM Grid Preparation at Different Timescales. *Structure* **28,** 1238-1248.e4 (2020).
292. Luoto, H. H., Baykov, A. A., Lahti, R. & Malinen, A. M. Membrane-integral pyrophosphatase subfamily capable of translocating both Na^+ and H^+. *PNAS* **110,** 1255–1260 (2013).
293. Liu, Y., Huynh, D. T. & Yeates, T. O. A 3.8 Å resolution cryo-EM structure of a small protein bound to an imaging scaffold. *Nat. Commun.* **10,** 1864 (2019).
294. Frank, G. A. *et al.* Cryo-EM Analysis of the Conformational Landscape of Human P-glycoprotein (ABCB1) During its Catalytic Cycle. *Mol. Pharmacol.* **90,** 35–41 (2016).
295. Schultz, A. & Baltscheffsky, M. Inhibition studies on *Rhodospirillum rubrum* H^+-pyrophosphatase expressed in *Escherichia coli*. *Biochim. Biophys. Acta Bioenerg.* **1656,** 156–165 (2004).

Appendix

Supplementary information Chapter 1

Table S1: Structural alignment of M-PPase structures.

Rmsd (Å)[Ŧ, *]	5LZQ	4AV6	5LZR	4AV3	6QXA	4A01	5GPJ	6AFS
5LZQ	0.32	0.82	1.41	1.21	0.39	0.86	-	-
4AV6	0.82	0.13	-	-	0.86	-	-	1.17
5LZR	1.41	-	0.40	0.40	1.34	-	1.12	-
4AV3	1.21	-	0.40	0.10	1.11	-	-	-
6QXA	0.39	0.86	1.34	1.11	0.32	-	-	0.86
4A01	0.86	-	-	-	-	0.14	0.78	0.11
5GPJ	-	-	1.12	-	-	0.78	0.31	0.76
6AFS	-	1.17	-	-	0.86	0.11	0.76	018

Ŧ Structural alignment based on Cα atoms of subunit A. Comparison of identical structures refers to structural alignment of subunit A and subunit B instead.
* PDB identifier used in row and column header. 5LZQ: Tm-PPase:Mg_5IDP, 4AV6: *Tm*-PPase:Mg_4P_{i2}, 5LZR: *Tm*-PPase:Mg_2WO_4, 4AV3: *Tm*-PPase:CaMg , 6QXA: *Tm*-PPase:ATC_2, 4A01: Vr-PPase:Mg_5IDP, 5GPJ: *Vr*-PPase:Mg_2P_i, 6AFS: *Vr*-PPase: Mg_5P_{i2}

Supplementary information Chapter 2

Table S2: List of primers

Primer †	Description	Sequence ‡	ID
pDDGFP_F	Amplification of pDDGFP2	CTGCAGGAATTCGATATCAAGCTTATC	#001
pDDGFP_R	Amplification of pDDGFP2	GGGGGATCCACTAGTTCTAGAATC	#002
syn_*Cl*-PPase_F	Testing presence of synthetic *Cl*-PPase mutagenesis construct	ACCCCGGATTCTAGAACTAG	#003
syn_*Cl*-PPase_R	Testing presence of synthetic *Cl*-PPase mutagenesis construct	ATCGATAAGCTTGATATCGAATTCC	#004
C544A_F	Introducing C544A in *Cp*-PPase sequence	GGTCGGCCGCgctGCCCAGAAGA	#005
C544A_R	Introducing C544A in *Cp*-PPase sequence	GCATCCATGGTGAGGGCGGC	#006
S113A_F	Introducing S113A in *Cp*-PPase sequence	CACCGCCAACgctGCCAGCAAAAGCCTGA	#007
S113A_R	Introducing S113A in *Cp*-PPase sequence	CGGCCGTTGGCATAGGTCGC	#008
S273V_F	Introducing S273V in *Cp*-PPase sequence	GATCCCCATGgttATGGCCGCTATTGGCATTCTGGCTTCC	#009
S273V_R	Introducing S273V in *Cp*-PPase sequence	ACCACGCCCTTAAAGCCTAAGC	#010
F334V_F	Introducing F334V in *Cp*-PPase sequence	AAAGATCGGCgtcTATTTCGCCATTCTTTCCG	#011
F334V_R	Introducing F334V in *Cp*-PPase sequence	TCCGCGCCCAGGCCGAAGT	#012
V350F_F	Introducing V350F in *Cp*-PPase sequence	TCTCATCGGTtttGTCACCGAATATTACACCTCCGATTCC	#013
V350F_R	Introducing V350F in *Cp*-PPase sequence	ATACCAGCCAGCAGGCCGG	#014
G108A_F	Introducing G108A in *Cp*-PPase sequence	CTATGCCAACgccCGCACCGCCAAC	#015
G108A_R	Introducing G108A in *Cp*-PPase sequence	GTCGCGATCTTCATGCCGATAAAGCCGG	#016
A499D_F	Introducing A499D in *Cp*-PPase sequence	TTCCTATATCgatGAAATCCAGGTGCTGAAGCCTGA	#017
A499D_R	Introducing A499D in *Cp*-PPase sequence	GCGATCAAGGCCAGGGCG	#018
A398I_F	Introducing A398I in *Cp*-PPase sequence	CATTGTCGGCattTCCGTTTTGGTCAGCTATTTTCTGTCC	#019
A398I_R	Introducing A398I in *Cp*-PPase sequence	ACTACAGGAAGCACGGTGGAAAGC	#020
T564E_F	Introducing T564E in *Cp*-PPase sequence	GGGACTGATGgaaGGGGAAGCGGATCCTGATTACGCTTCC	#021

T564E_R	Introducing T564E in *Cp*-PPase sequence	TTGATCTCACGGAACTGGCGG	#022
L391I_F	Introducing L391I in *Cp*-PPase sequence	TTCCACCGTGattCCTGTAGTCATTGTCGG	#023
L391I_R	Introducing L391I in *Cp*-PPase sequence	AGCATTCCCAGGGACAGGC	#024
S22A_F	Introducing S22A in *Cp*-PPase sequence	GTTTTTCCTTgccCGCAGAGTTATGAAGGC	#025
S22A_R	Introducing S22A in *Cp*-PPase sequence	GCGAACAGCAGAGCCACTACC	#026
A429S_F	Introducing A429S in *Cp*-PPase sequence	CGGTATGCTGagtACCCTGGGCATTACCCTGG	#027
A429S_R	Introducing A429S in *Cp*-PPase sequence	ACGGCGGACAGGCCG	#028
I124V_F	Introducing I124V in *Cp*-PPase sequence	GGGCCTGAGAgtcGCTTTCTCCG	#029
I124V_R	Introducing I124V in *Cp*-PPase sequence	GCGTTCAGGCTTTTGCTGGC	#030
G130S_F	Introducing G130S in *Cp*-PPase sequence	CTCCGCCGGCagcGTAATGGGCT	#031
G130S/A_R	Introducing G130S in *Cp*-PPase sequence	AAAGCGATTCTCAGGCCCGCGTTC	#032
G326L_F	Introducing G326L in *Cp*-PPase sequence	GATTTACTTCttgCTGGGCGCGGAAAAGATCGGCT	#033
G326L_R	Introducing G326L in *Cp*-PPase sequence	AGCGGGAAAGCGGCCACAG	#034
S371K_F	Introducing S371K in *Cp*-PPase sequence	GGGCACCTCCaaaACCGGCCCCG	#035
S371K_R	Introducing S371K in *Cp*-PPase sequence	GCTAGCTTCTTGGTAGGCTTATAGGAAT	#036
A329P_F	Introducing A329P in *Cp*-PPase sequence	CGGCCTGGGCccaGAAAAGATCGGC	#037
A329P_R	Introducing A329P in *Cp*-PPase sequence	AAGTAAATCAGCGGGAAAGCGG	#038
V83F_F	Introducing V83F in *Cp*-PPase sequence	CTTTGTGCCGtttGCGTTCCTGACC	#039
V83F_R	Introducing V83F in *Cp*-PPase sequence	GGGGTCAGGAACCCGGC	#040
V351F_F	Introducing V351F in *Cp*-PPase sequence	CATCGGTGTGttcACCGAATATTACACCTCCG	#041
V351F_R	Introducing V351F in *Cp*-PPase sequence	AGAATACCAGCCAGCAGGCC	#042
W309Y_F	Introducing W309Y in *Cp*-PPase sequence	GCTTGGAACCtatATTTCCGCCGCTTTGATCGCTGTG	#043
W309Y_R	Introducing W309Y in *Cp*-PPase sequence	CGCAGAGATCCCAGAAGGTTTTTCT	#044
K331H_F	Introducing K331H in *Cp*-PPase sequence	GGGCGCGGAAcatATCGGCTTCTATTTCGCCATTC	#045
K331H_R	Introducing K331H	AGGCCGAAGTAAATCAGCGGG	#046
F20Y_F	Introducing F20Y in *Cp*-PPase sequence	GTTCGCGTTTtacCTTTCCCGCAGAGTTATGAA	#047
F20Y_R	Introducing F20Y in *Cp*-PPase sequence	AGCAGAGCCACTACCGCG	#048
S303A_F	Introducing S303A in *Cp*-PPase sequence	CCTTCTGGGAgctCTGCGGCTTG	#049
S303A_R	Introducing S303A in *Cp*-PPase sequence	TTTTTCTGAGAGGCGTTTTCCTTGGT	#050
M456L_F	Introducing M456L in *Cp*-PPase sequence	GATGGCCCACttgGGCGAGGAGG	#051
M456L_R	Introducing M456L in *Cp*-PPase sequence	TCCGCGATACCGCCGGCGTTAT	#052
V269A_F	Introducing V269A in *Cp*-PPase sequence	TAAGGGCGTGgctATCCCCATGTCCA	#053
V269A_R	Introducing V269A in *Cp*-PPase sequence	AAGCCTAAGCCGGCGG	#054
R109W_F	Introducing R109W in *Cp*-PPase sequence	TGCCAACGGCtggACCGCCAACA	#055
R109W_R	Introducing R109W in *Cp*-PPase sequence	TAGGTCGCGATCTTCATGCCGATA	#056
V81W_F	Introducing V81W in *Cp*-PPase sequence	GACCCCCTTTtggCCGGTTGCGTTC	#057
V81W_R	Introducing V81W in *Cp*-PPase sequence	AGGAACCCGGCGATGGC	#058
I323W_F	Introducing I323W in *Cp*-PPase sequence	TTTCCCGCTGtggTACTTCGGCCTGGGC	#059

I323W_R	Introducing I323W in *Cp*-PPase sequence	GCGGCCACAGCGATCAAAGCG	#060
D468F_F	Introducing D468F in *Cp*-PPase sequence	CGACGCTTTGtttTCCCTGGGCAACACCACT	#061
D468F_R	Introducing D468F in *Cp*-PPase sequence	GTGCGCTTGCGGACCT	#062
F80I_F	Introducing F80I in *Cp*-PPase sequence	CCTGACCCCCattGTGCCGGTTG	#063
F80I_R	Introducing F80 in *Cp*-PPase sequence I	AACCCGGCGATGGCC	#064
Q162Y_F	Introducing Q162Y in *Cp*-PPase sequence	CGACGTGATGtatAACGAGGCCGCCCAGG	#065
Q162Y_R	Introducing Q162Y in *Cp*-PPase sequence	ACAGTGCTGTACCAGAACTTCAGG	#066
N509Y_F	Introducing N509Y in *Cp*-PPase sequence	GCCTGATTTCtacTTTAACCTGACCATTACCAATCCCCCC	#067
N509Y_R	Introducing N509Y in *Cp*-PPase sequence	TTCAGCACCTGGATTTCCGCGATAT	#068
L151I_F	Introducing L151I in *Cp*-PPase sequence	GTATTTCTTCattAAGTTCTGGTACAGCACTGTCGACGTG	#069
L151I_R	Introducing L151I in *Cp*-PPase sequence	CAGATAGAGATATCCAGCAGGCC	#070
F61W_F	Introducing F61W in *Cp*-PPase sequence	GGCTATTTTCtggGCGGTTATGTTCGTTATCCTGCTGATC	#071
F61W_R	Introducing F61W in *Cp*-PPase sequence	ACGCCGGTATACTGGCGCTTC	#072
P82W_F	Introducing P82W in *Cp*-PPase sequence	CCCCTTTGTGtggGTTGCGTTCCTGACCGGAGGC	#073
P82W_R	Introducing P82W in *Cp*-PPase sequence	GTCAGGAACCCGGCGATGGC	#074
R463W_F	Introducing R463W in *Cp*-PPase sequence	GGTCCGCAAGtggACCGACGCTTTGGATTC	#075
R463W_R	Introducing R463W in *Cp*-PPase sequence	TCCTCGCCCATGTGGGC	#076
I501L_F	Introducing I501L in *Cp*-PPase sequence	TATCGCGGAAttgCAGGTGCTGAAGCCTGA	#077
I501L_R	Introducing I501L in *Cp*-PPase sequence	TAGGAAGCGATCAAGGCCAG	#078
D358W_F	Introducing D358W in *Cp*-PPase sequence	TTACACCTCCtggTCCTATAAGCCTACCAAGAAGCTAGCG	#079
D358W_R	Introducing D358W in *Cp*-PPase sequence	TATTCGGTGACCACACCGATGAG	#080
I171Y_F	Introducing I171Y in *Cp*-PPase sequence	GGTGCAGGCGtatACTTCCGCCATGCTTACCTTTG	#081
I171Y_R	Introducing I171Y in *Cp*-PPase sequence	TGGGCGGCCTCGTTCT	#082
R290F_F	Introducing R290F in *Cp*-PPase sequence	CTTCTTTGTAttcACCAAGGAAAACGCCTCTCAGAAAAAC	#083
R290F_R	Introducing R290F in *Cp*-PPase sequence	GTGCCGATGATGGAAGCCAG	#084
Y418S_F	Introducing Y418S in *Cp*-PPase sequence	TAACGGCCTCagcGGCGTCGGCCTGTCCGC	#085
Y418S_R	Introducing Y418S in *Cp*-PPase sequence	TTATAGGAGTCGCTGCCGCCGGACAGA	#086
L69Y_F	Introducing L69Y in *Cp*-PPase sequence	CGTTATCCTGtatATCCTGGCCATCGCCGGGTTCCTGACC	#087
L69Y_R	Introducing L69Y in *Cp*-PPase sequence	AACATAACCGCGAAGAAAATAGCCA	#088
V693Y_F	Introducing V693Y in *Cp*-PPase sequence	CGCCAGCGTGtatCTTTCCTTCAGCCTGA	#089
V693Y_R	Introducing V693Y in *Cp*-PPase sequence	AACACGATGGAAACCATGCTCAGC	#090
T363W_F	Introducing T363W in *Cp*-PPase sequence	CTATAAGCCTtggAAGAAGCTAGCGGGCACCTCCAGCACC	#091
T363W_R	Introducing T363W in *Cp*-PPase sequence	GAATCGGAGGTGTAATATTCGGTGAC	#092
L435W_F	Introducing L435W in *Cp*-PPase sequence	GGGCATTACCtggGCCACCGACGCTTA	#093
L435W_R	Introducing L435W in *Cp*-PPase sequence	AGGGTAGCCAGCATACCGAC	#094
I73P_F	Introducing I73P in *Cp*-PPase sequence	GATCCTGGCCccaGCCGGGTTCCTGAC	#095
I73P_R	Introducing I73P in *Cp*-PPase sequence	AGCAGGATAACGAACATAACCGC	#096
L142P_F	Introducing L142P in *Cp*-PPase sequence	CCTGGGCCTGccaGATATCTCTATCTGGTATTTCTTCCTG	#097

L142P_R	Introducing L142P in *Cp*-PPase sequence	CCTACCACCACAAAGCCCATTACG	#098
V519D_F	Introducing V519D in *Cp*-PPase sequence	CAATCCCCCCgatCTCATCGGCCTGTTCA T	#099
V519D_R	Introducing V519D in *Cp*-PPase sequence	GTAATGGTCAGGTTAAAGTTGAAATCAG	#100
L422W_F	Introducing L422W in *Cp*-PPase sequence	CGGCGTCGGCtggTCCGCCGTC	#101
L422W_R	Introducing L422W in *Cp*-PPase sequence	TAGAGGCCGTTATTATAGGAGTCGCTGC	#102
I433W_F	Introducing I433W in *Cp*-PPase sequence	TACCCTGGGCtggACCCTGGCCACCGAC GCTTAC	#103
I433W_R	Introducing I433W in *Cp*-PPase sequence	GCCAGCATACCGACGGCGG	#104
G179A_F	Introducing G179A in *Cp*-PPase sequence	GCTTACCTTTgctATGGGCGCTTCCTCTAT G	#105
G179A_R	Introducing G179A in *Cp*-PPase sequence	ATGGCGGAAGTAATCGCCTG	#106
T430A_F	Introducing T430A in *Cp*-PPase sequence	TATGCTGGCTgccCTGGGCATTA	#107
T430A_R	Introducing T430A in *Cp*-PPase sequence	CCGACGGCGGACAGGC	#108
A275L_F	Introducing A275L in *Cp*-PPase sequence	CATGTCCATGttgGCTATTGGCATTCTGGC TTCCATCAT	#109
A275L_R	Introducing A275L in *Cp*-PPase sequence	GGGGATCACCACGCCCTT	#110
A492L_F	Introducing A492L in *Cp*-PPase sequence	CACCGCCCTGttgTTGATCGCTTCCTATAT CGCGGA	#111
A492L_R	Introducing A492L in *Cp*-PPase sequence	AGAGCGGCGGAGCCG	#112
G526A_F	Introducing G526A in *Cp*-PPase sequence	CCTGTTCATCgccGGCGTGCTGC	#113
G526A_R	Introducing G526A in *Cp*-PPase sequence	CCGATGAGTACGGGGGGATTGGT	#114
G88A_F	Introducing G88A in *Cp*-PPase sequence	GTTCCTGACCgcaGGCTTCTTCTCC	#115
G88A_R	Introducing G88A in *Cp*-PPase sequence	GCAACCGGCACAAAGGGG	#116
G620A_F	Introducing G620A in *Cp*-PPase sequence	CACAGTCTCCgcaTTTATTCTCGCCGTTAT GATG	#117
G620A_R	Introducing G620A in *Cp*-PPase sequence	GCGCCGGCCAGCAT	#118
G381A_F	Introducing G381A in *Cp*-PPase sequence	CATCATCGGCgccCTGTCCCTGG	#119
G381A_R	Introducing G381A in *Cp*-PPase sequence	ACGGTAGCGGGGCCGG	#120
S95A_F	Introducing S95A in *Cp*-PPase sequence in *Cp*-PPase sequence	CTCCGCCTTGgccGGCTTTATCG	#121
S95A_R	Introducing S95A in *Cp*-PPase sequence	AAGAAGCCTCCGGTCAGGAACG	#122
S184A_F	Introducing S184A in *Cp*-PPase sequence	GGGCGCTTCCgctATGGCTCTGT	#123
S184A_R	Introducing S184A in *Cp*-PPase sequence	ATACCAAAGGTAAGCATGGCGGAAGTAA	#124
P321A_F	Introducing P321A in *Cp*-PPase sequence	GGCCGCTTTCgctCTGATTTACTTCGGCC TG	#125
P321A_R	Introducing P321A in *Cp*-PPase sequence	ACAGCGATCAAAGCGGCG	#126
A281L_F	Introducing A281L in *Cp*-PPase sequence	TGGCATTCTGttgTCCATCATCGGCACCTT CTTTGTACGC	#127
A281L_R	Introducing A281L in *Cp*-PPase sequence	ATAGCGGCCATGGACATGGG	#128
A93L_F	Introducing A93L in *Cp*-PPase sequence	CTTCTTCTCCttgTTGTCCGGCTTTATCGG CATGAAGATC	#129
A93L_R	Introducing A93L in *Cp*-PPase sequence	CCTCCGGTCAGGAACGCAAC	#130
A312L_F	Introducing A312L in *Cp*-PPase sequence	CTGGATTTCCttgGCTTTGATCGCTGTGGC CGCTTTCCC	#131
A312L_R	Introducing A312L in *Cp*-PPase sequence	GTTCCAAGCCGCAGAGATCCCAG	#132
A624L_F	Introducing A624L in *Cp*-PPase sequence	ATTTATTCTCttgGTTATGATGGCAAACTC CGGCGGCGCC	#133
A624L_R	Introducing A624L in *Cp*-PPase sequence	CCGGAGACTGTGGCGCCG	#134

A84L_F	Introducing A84L in *Cp*-PPase sequence	TGTGCCGGTTttgTTCCTGACCGGAGG	#135
A84L_R	Introducing A84L in *Cp*-PPase sequence	AAGGGGGTCAGGAACCCG	#136
A313L_F	Introducing A313L in *Cp*-PPase sequence	GATTTCCGCCttgTTGATCGCTGTGGCCG CTTTCCC	#137
A313L_R	Introducing A313L in *Cp*-PPase sequence	CAGGTTCCAAGCCGCAGAGATC	#138
A616L_F	Introducing A616L in *Cp*-PPase sequence	GCTGGCCGGCttgACAGTCTCCGGATTTA TTCTCGCCGTT	#139
A616L_R	Introducing A616L in *Cp*-PPase sequence	ATACCGGCAACGCCGTTTACGC	#140
G527A_F	Introducing G527A in *Cp*-PPase sequence	GTTCATCGGCgccGTGCTGCCGT	#141
G527A_R	Introducing G527A in *Cp*-PPase sequence	AGGCCGATGAGTACGGGGGGATTG	#142
G140A_F	Introducing G140A in *Cp*-PPase sequence	GGTAGGCCTGgccCTGCTGGATATCTCTA TCTGGTATTTC	#143
G140A_R	Introducing G140A in *Cp*-PPase sequence	ACCACAAAGCCCATTACGCCGC	#144
G31A_F	Introducing G31A in *Cp*-PPase sequence	GGCTGACGAGgccACCGACCTGA	#145
G31A_R	Introducing G31A in *Cp*-PPase sequence	TTCATAACTCTGCGGGAAAGGAAAAACG	#146
G368A_F	Introducing G368A in *Cp*-PPase sequence	GAAGCTAGCGgccACCTCCAGCA	#147
G368A_R	Introducing G368A in *Cp*-PPase sequence	TTGGTAGGCTTATAGGAATCGGAGGTGT	#148
G302A_F	Introducing G302A in *Cp*-PPase sequence	AAACCTTCTGgcaTCTCTGCGGCTT	#149
G302A_R	Introducing G302A in *Cp*-PPase sequence	TTCTGAGAGGCGTTTTCCTTGGT	#150
A319L_F	Introducing A319L in *Cp*-PPase sequence	CGCTGTGGCCttgTTCCCGCTGATTTACTT CGGCCTG	#151
A319L_R	Introducing A319L in *Cp*-PPase sequence	ATCAAAGCGGCGGAAATCCAGG	#152
E30A_F	Introducing E30A in *Cp*-PPase sequence	GAAGGCTGACgctGGCACCGACCTGATGA AGAAAATC	#153
E30A_R	Introducing E30A in *Cp*-PPase sequence	ATAACTCTGCGGGAAAGGAAAAACGC	#154
E458A_F	Introducing E458A in *Cp*-PPase sequence	CCACATGGGCgctGAGGTCCGCA	#155
E458A_R	Introducing E458A in *Cp*-PPase sequence	GCCATCTCCGCGATACCGC	#156
E293A_F	Introducing E293A in *Cp*-PPase sequence	ACGCACCAAGgcaAACGCCTCTC	#157
E293A_R	Introducing E293A in *Cp*-PPase sequence	ACAAAGAAGGTGCCGATGATGGAAG	#158
G672A_F	Introducing G672A in *Cp*-PPase sequence	GGATACCTCCgccCCGTCTATCAACA	#159
G672A_R	Introducing G672A in *Cp*-PPase sequence	TTGAAGGGATCGCCCACG	#160
G230A_F	Introducing G230A in *Cp*-PPase sequence	CGATAACGTAgccGACAACGTAGGCGA	#161
G230A_R	Introducing G230A in *Cp*-PPase sequence	GCGATAACGGCGGGGTTG	#162
G238A_F	Introducing G238A in *Cp*-PPase sequence	CGACGTTGCCgccATGGGCGCCGACCTG TACG	#163
G238A_R	Introducing G238A in *Cp*-PPase sequence	CCTACGTTGTCGCCTACGTTATCGGCGA	#164
G234A_F	Introducing G234A in *Cp*-PPase sequence	CGACAACGTAgccGACGTTGCCG	#165
G234A_R	Introducing G234A in *Cp*-PPase sequence	CCTACGTTATCGGCGATAACGGC	#166
A316L_F	Introducing A316L in *Cp*-PPase sequence	CGCTTTGATCttgGTGGCCGCTTTCCCGC TGA	#167
A316L_R	Introducing A316L in *Cp*-PPase sequence	GCGGAAATCCAGGTTCCAAGCC	#168
S619A_F	Introducing S619A in *Cp*-PPase sequence	CGCCACAGTCgccGGATTTATTCTCGCCG TTAT	#169
S619A_R	Introducing S619A in *Cp*-PPase sequence	CCGGCCAGCATACCGGCAAC	#170
S340A_F	Introducing S340A in *Cp*-PPase sequence	CGCCATTCTTgccGGCCTGCTGG	#171
S340A_R	Introducing S340A in *Cp*-PPase sequence	AAATAGAAGCCGATCTTTTCCGCGCCC	#172

F531A_F	Introducing F531A in *Cp*-PPase sequence	CGTGCTGCCGgctCTGTTCGCCGC	#173
F531A_R	Introducing F531A in *Cp*-PPase sequence	CCG ATG AAC AGG CCG ATG	#174
E642A_F	Introducing E642A in *Cp*-PPase sequence	GAAGTACATTgcaAGCGGCGAGTATGGC	#175
E642A_R	Introducing E642A in *Cp*-PPase sequence	TTAGCGTTATCCCAGGCGCC	#176
A690L_F	Introducing A690L in *Cp*-PPase sequence	CATCGTGTTCttgAGCGTGGTGCTTTCCTTCAGCCTGAT	#177
A690L_R	Introducing A690L in *Cp*-PPase sequence	GAAACCATGCTCAGCAGCTTG	#178
G96A_F	Introducing G96A in *Cp*-PPase sequence	CGCCTTGTCCgccTTTATCGGCAT	#179
G96A_R	Introducing G96A in *Cp*-PPase sequence	GAGAAGAAGCCTCCGGTCAGG	#180
A14L_F	Introducing A14L in *Cp*-PPase sequence	CGCGGTAGTGttgCTGCTGTTCGCGTTTTTCCTTTCCCG	#181
A14L_R	Introducing A14L in *Cp*-PPase sequence	CCGACAGGAGCAAGGATGGAC	#182
G480A_F	Introducing G480A in *Cp*-PPase sequence	TACCGGCAAGgccTTTGCCATCG	#183
G480A_R	Introducing G480A in *Cp*-PPase sequence	GCGGCAGTGGTGTTGCCC	#184
A535L_F	Introducing A535L in *Cp*-PPase sequence	CCTGTTCGCCttgCTCACCATGGATGCGGTCGG	#185
A535L_R	Introducing A535L in *Cp*-PPase sequence	AACGGCAGCACGCCGCCG	#186
A495L_F	Introducing A495L in *Cp*-PPase sequence	GGCCTTGATCttgTCCTATATCGCGGAAATCCAGGTGCTG	#187
A495L_R	Introducing A495L in *Cp*-PPase sequence	AGGGCGGTGAGAGCGGCG	#188
G397A_F	Introducing G397A in *Cp*-PPase sequence	AGTCATTGTCgccGCTTCCGTTTTGG	#189
G397A_R	Introducing G397A in *Cp*-PPase sequence	ACAGGAAGCACGGTGGAAAGC	#190
T670A_F	Introducing T670A in *Cp*-PPase sequence	CTTCAAGGATgccTCCGGCCCGTCTAT	#191
T670A_R	Introducing T670A in *Cp*-PPase sequence	GGATCGCCCACGGTATCGCC	#192
K292A_F	Introducing K292A in *Cp*-PPase sequence	TGTACGCACCgctGAAAACGCCTCTCAGAAAAACCTTCTG	#193
K292A_R	Introducing K292A in *Cp*-PPase sequence	AAGAAGGTGCCGATGATGGAAGC	#194
A114L_F	Introducing A114L in *Cp*-PPase sequence	CGCCAACAGCttgAGCAAAAGCCTGAACGCGGG	#195
A114L_R	Introducing A114L in *Cp*-PPase sequence	GTGCGGCCGTTGGCATAGG	#196
G130A_F	Introducing G130A in *Cp*-PPase sequence	CTCCGCCGGCGcCGTAATGGGCT	#197
A344S_F	Introducing A344S in *Cp*-PPase sequence	GGCCTGCTGtCTGGTATTCTCATCGGTGT	#198
A344S_R	Introducing A344S in *Cp*-PPase sequence	GGAAAGAATGGCGAAATAGAAGCC	#199

† Primer direction indicated by F for forward or R for reverse at end of the primer name

‡ Lower case letters indicate the introduction of point mutations

Table S3: List of all construct sequences used in this study in fasta format

Construct sequences
>pDDGFP2
GACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGTATGATCCAAT
ATCAAAGGAAATGATAGCATTGAAGGATGAGACTAATCCAATTGAGGAGTGGCAGCATATAGAACAGCTAAAGG
GTAGTGCTGAAGGAAGCATACGATACCCCGCATGGAATGGGATAATATCACAGGAGGTACTAGACTACCTTTCA
TCCTACATAAATAGACGCATATAAGTACGCATTTAAGCATAAACACGCACTATGCCGTTCTTCTCATGTATATATA
TATACAGGCAACACGCAGATATAGGTGCGACGTGAACAGTGAGCTGTATGTGCGCAGCTCGCGTTGCATTTTCG
GAAGCGCTCGTTTTCGGAAACGCTTTGAAGTTCCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCA
GAGCGCTTTTGAAAACCAAAAGCGCTCTGAAGACGCACTTTCAAAAAACCAAAAACGCACCGGACTGTAACGAG
CTACTAAAATATTGCGAATACCGCTTCCACAAACATTGCTCAAAAGTATCTCTTTGCTATATATCTCTGTGCTATAT
CCCTATATAACCTACCCATCCACCTTTCGCTCCTTGAACTTGCATCTAAACTCGACCTCTACATTTTTTATGTTTAT
CTCTAGTATTACTCTTTAGACAAAAAAATTGTAGTAAGAACTATTCATAGAGTGAATCGAAAACAATACGAAAATG
TAAACATTTCCTATACGTAGTATATAGAGACAAAATAGAAGAAACCGTTCATAATTTTCTGACCAATGAAGAATCA
TCAACGCTATCACTTTCTGTTCACAAAGTATGCGCAATCCACATCGGTATAGAATATAATCGGGGATGCCTTTAT
CTTGAAAAAATGCACCCGCAGCTTCGCTAGTAATCAGTAAACGCGGGAAGTGGAGTCAGGCTTTTTTTATGGAA

GAGAAAATAGACACCAAAGTAGCCTTCTTCTAACCTTAACGGACCTACAGTGCAAAAAGTTATCAAGAGACTGCA
TTATAGAGCGCACAAAGGAGAAAAAAAGTAATCTAAGATGCTTTGTTAGAAAAATAGCGCTCTCGGGATGCATTT
TTGTAGAACAAAAAAGAAGTATAGATTCTTTGTTGGTAAAATAGCGCTCTCGCGTTGCATTTCTGTTCTGTAAAAA
TGCAGCTCAGATTCTTTGTTTGAAAAATTAGCGCTCTCGCGTTGCATTTTTGTTTTACAAAAATGAAGCACAGATT
CTTCGTTGGTAAAATAGCGCTTTCGCGTTGCATTTCTGTTCTGTAAAAATGCAGCTCAGATTCTTTGTTTGAAAAA
TTAGCGCTCTCGCGTTGCATTTTTGTTCTACAAAATGAAGCACAGATGCTTCGTTCAGGTGGCACTTTTCGGGGA
AATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCT
GATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTT
TGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGG
GTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGT
TTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAA
CTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGAT
GGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGAC
AACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTT
GGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAAC
GTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGG
ATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGT
GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACA
CGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCA
TTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTA
GGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCC
CGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCA
CCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAG
AGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGC
CTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTG
GACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGC
TTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAG
GGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGG
GGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCG
TCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTT
TTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATA
CCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGC
AAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGG
GCAGTGAGCGCAACGCAATTAATGTGAGTTACCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCG
GCTCCTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAG
CGCGCAATTAACCCTCACTAAAGGGAACAAAAGCTGGAGCTCACGGATTAGAAGCCGCCGAGCGGGTGACAGC
CCTCCGAAGGAAGACTCTCCTCCGTGCGTCCTCGTCTTCACCGGTCGCGTTCCTGAAACGCAGATGTGCCTCG
CGCCGCACTGCTCCGAACAATAAAGATTCTACAATACTAGCTTTTATGGTTATGAAGAGGAAAAATTGGCAGTAA
CCTGGCCCCACAAACCTTCAAATGAACGAATCAAATTAACAACCATAGGATGATAATGCGATTAGTTTTTTAGCCT
TATTTCTGGGGTAATTAATCAGCGAAGCGATGATTTTTGATCTATTAACAGATATATAAATGCAAAAACTGCATAA
CCACTTTAACTAATACTTTCAACATTTTCGGTTTGTATTACTTCTTATTCAAATGTAATAAAAGTATCAACAAAAAAT
TGTTAATATACCTCTATACTTTAACGTCAAGGAGAAAAAACCCCGGATTCTAGAACTAGTGGATCCCCCCCCGGG
GAAAATTTATATTTTCAAGGTCAATTTTCTAAAGGTGAAGAATTATTCACTGGCGTTGTCCCAATTTTGGTTGAATT
AGATGGTGATGTTAATGGTCACAAATTTTCTGTCTCCGGTGAAGGTGAAGGTGACGCTACTTACGGTAAATTGAC
CTTAAAATTTATTTGTACTACTGGTAAATTGCCAGTTCCATGGCCAACCTTAGTCACTACTTTCGGTTATGGTGTT
CAATGTTTTGCGAGATACCCAGATCATATGAAACAACATGACTTTTTCAAGTCTGCCATGCCAGAAGGTTATGTT
CAAGAAAGAACTATTTTTTTCAAAGATGACGGTAACTACAAGACCAGAGCTGAAGTCAAGTTTGAAGGTGATACC
TTAGTTAATAGAATCGAATTAAAAGGTATTGATTTTAAAGAAGATGGTAACATTTTAGGTCACAAATTGGAATACA
ACTATAACTCTCACAATGTTTACATCATGGCTGACAAACAAAAGAATGGTATCAAAGTTAACTTCAAAATTAGACA
CAACATTGAAGATGGTTCTGTTCAATTAGCTGACCATTATCAACAAAATACTCCAATTGGTGATGGTCCAGTCTTG
TTACCAGACAACCATTACTTATCCACTCAATCTGCCTTATCCAAAGATCCAAACGAAAAGAGAGACCACATGGTC
TTGTTAGAATTTGTTACTGCTGCTGGTATTACCCATGGTATGGATGAATTGTACAAACACCACCACCATCATCATC
ATCATTAACTGCAGGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGTCATGTAATTAGTTATGTCACG
CTTACATTCACGCCCCTCCCCCCACATCCGCTCTAACCGAAAAGGGAAGGAGTTAGACAACCTGAAGTCTAGGT
CCCCTATTTATTTTTTTATAGTTATGTTAGTATTAAGAACGTTATTTATATTCAAATTTTCCTTTTTTTTTCTGTACAGA
CGCGTGTACGCATGTAACATTATACTGAAAACCTTGCTTGAGAAGGTTTTGGGACCCAATTCGCCCTATAGTGAG
TCGTATTACGCGCGCTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAA
TCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAA
CAGTTGCGCAGCCTGAATGGCGAATGGCGCGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTG
GTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCT
CGCCACGTTCGCCGGCAAATGGTCAGGTCATTGAGTGTTTTTTATTTGTTGTATTTTTTTTTTTTTTAGAGAAAATC
CTCCAATATCAAATTAGGAATCGTAGTTTCATGATTTTCTGTTACACCTAACTTTTTGTGTGGTGCCCTCCTCCTT
GTCAATATTAATGTTAAAGTGCAATTCTTTTTCCTTATCACGTTGAGCCATTAGTATCAATTTGCTTACCTGTATTC
CTTTACTATCCTCCTTTTTCTCCTTCTTGATAAATGTATGTAGATTGCGTATATAGTTTCGTCTACCCTATGAACAT
ATTCCATTTTGTAATTTCGTGTCGTTTCTATTATGAATTTCATTTATAAAGTTTATGTACAAATATCATAAAAAAAGA
GAATCTTTTTAAGCAAGGATTTTCTTAACTTCTTCGGCGACAGCATCACCGACTTCGGTGGTACTGTTGGAACCA
CCTAAATCACCAGTTCTGATACCTGCATCCAAAACCTTTTTAACTGCATCTTCAATGGCCTTACCTTCTTCAGGCA
AGTTCAATGACAATTTCAACATCATTGCAGCAGACAAGATAGTGGCGATAGGGTCAACCTTATTCTTTGGCAAAT
CTGGAGCAGAACCGTGGCATGGTTCGTACAAACCAAATGCGGTGTTCTTGTCTGGCAAAGAGGCCAAGGACGC
AGATGGCAACAAACCCAAGGAACCTGGGATAACGGAGGCTTCATCGGAGATGATATCACCAAACATGTTGCTGG
TGATTATAATACCATTTAGGTGGGTTGGGTTCTTAACTAGGATCATGGCGGCAGAATCAATCAATTGATGTTGAA
CCTTCAATGTAGGGAATTCGTTCTTGATGGTTTCCTCCACAGTTTTTCTCCATAATCTTGAAGAGGCCAAAAGATT
AGCTTTATCCAAGGACCAAATAGGCAATGGTGGCTCATGTTGTAGGGCCATGAAAGCGGCCATTCTTGTGATTC
TTTGCACTTCTGGAACGGTGTATTGTTCACTATCCCAAGCGACACCATCACCATCGTCTTCCTTTCTCTTACCAAA
GTAAATACCTCCCACTAATTCTCTGACAACAACGAAGTCAGTACCTTTAGCAAATTGTGGCTTGATTGGAGATAA
GTCTAAAAGAGAGTCGGATGCAAAGTTACATGGTCTTAAGTTGGCGTACAATTGAAGTTCTTTACGGATTTTTAG
TAAACCTTGTTCAGGTCTAACACTACCGGTACCCCATTTAGGACCAGCCACAGCACCTAACAAAACGGCATCAA
CCTTCTTGGAGGCTTCCAGCGCCTCATCTGGAAGTGGGACACCTGTAGCATCGATAGCAGCACCACCAATTAAA

TGATTTTCGAAATCGAACTTGACATTGGAACGAACATCAGAAATAGCTTTAAGAACCTTAATGGCTTCGGCTGTG
ATTTCTTGACCAACGTGGTCACCTGGCAAAACGACGATCTTCTTAGGGGCAGACATTACAATGGTATATCCTTGA
AATATATATAAAAAAAGGCGCCTTAGACCGCTCGGCCAAACAACGCCGGCTTTCCCCGTCAAGCTCTAAATCGG
GGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTC
ACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGAC
TCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTC
GGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTT
CCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATAGGGTAATAACTGATATAATTAAAT
TGAAGCTCTAATTTGTGAGTTTAGTATACATGCATTTACTTATAATACAGTTTTTTAGTTTTGCTGGCCGCATCTTC
TCAAATATGCTTCCCAGCCTGCTTTTCTGTAACGTTCACCCTCTACCTTAGCATCCCTTCCCTTTGCAAATAGTCC
TCTTCCAACAATAATAATGTCAGATCCTGTAGAGACCACATCATCCACGGTTCTATACTGTTGACCCAATGCGTC
TCCCTTGTCATCTAAACCCACACCGGGTGTCATAATCAACCAATCGTAACCTTCATCTCTTCCACCCATGTCTCTT
TGAGCAATAAAGCCGATAACAAAATCTTTGTCGCTCTTCGCAATGTCAACAGTACCCTTAGTATATTCTCCAGTAG
ATAGGGAGCCCTTGCATGACAATTCTGCTAACATCAAAAGGCCTCTAGGTTCCTTTGTTACTTCTTCTGCCGCCT
GCTTCAAACCGCTAACAATACCTGGGCCCACCACACCGTGTGCATTCGTAATGTCTGCCCATTCTGCTATTCTGT
ATACACCCGCAGAGTACTGCAATTTGACTGTATTACCAATGTCAGCAAATTTTCTGTCTTCGAAGAGTAAAAAATT
GTACTTGGCGGATAATGCCTTTAGCGGCTTAACTGTGCCCTCCATGGAAAAATCAGTCAAGATATCCACATGTGT
TTTTAGTAAACAAATTTTGGGACCTAATGCTTCAACTAACTCCAGTAATTCCTTGGTGGTACGAACATCCAATGAA
GCACACAAGTTTGTTTGCTTTTCGTGCATGATATTAAATAGCTTGGCAGCAACAGGACTAGGATGAGTAGCAGCA
CGTTCCTTATATGTAGCTTTCGACATGATTTATCTTCGTTTCCTGCAGGTTTTTGTTCTGTGCAGTTGGGTTAAGA
ATACTGGGCAATTTCATGTTTCTTCAACACTACATATGCGTATATATACCAATCTAAGTCTGTGCTCCTTCCTTCG
TTCTTCCTTCTGTTCGGAGATTACCGAATCAAAAAAATTTCAAAGAAACCGAAATCAAAAAAAAGAATAAAAAAAA
AATGATGAATTGAATTGAAAAGCTGTGGTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGC
CAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGA
CAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGA

>syn-*Cl*-PPase

ACCCCGGATTCTAGAACTAGTGGATCCCCCATGCACCATCACCATCACCATCACCATAGCAAAGGAGAAGAACT
TTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCCGTGGAGA
GGGTGAAGGTGATGCTACAAACGGAAAACTCACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATG
GCCAACACTTGTCACTACTCTGACCTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGA
CTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGGACCTACAA
GACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAAGGTATTGATTTTAAAGA
AGATGGAAACATTCTCGGACACAAACTCGAGTACAACTTTAACTCACACAATGTATACATCACGGCAGACAAACA
AAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACGTTGAAGATGGTTCCGTTCAACTAGCAGACCATTA
TCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGTCCT
TTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACATG
GCATGGATGAGCTCTACAAGACCTCTGCTGGTGAGAACTTGTACTTTCAAGGTACTTCTGCTGGTGACCTGTCC
ATCCTTGCTCCTGTCGGCGCGGTAGTGGCTCTGCTGTTCGCGTTTTTCCTTTCCCGCAGAGTTATGAAGGCTGA
CGAGGGCACCGACCTGATGAAGAAAATCGCCGGCGCTGTCCGGAAAGGCGCCAACGCTTATCTGAAGCGCCA
GTATACCGGCGTGGCTATTTTCTTCGCGGTTATGTTCGTTATCCTGCTGATCCTGGCCATCGCCGGGTTCCTGA
CCCCCTTTGTGCCGGTTGCGTTCCTGACCGGAGGCTTCTTCTCCGCCTTGTCCGGCTTTATCGGCATGAAGATC
GCGACCTATGCCAACGGCCGCACCGCCAACAGCGCCAGCAAAAGCCTGAACGCGGGCCTGAGAATCGCTTTCT
CCGCCGGCGGCGTAATGGGCTTTGTGGTGGTAGGCCTGGGCCTGCTGGATATCTCTATCTGGTATTTCTTCCTG
AAGTTCTGGTACAGCACTGTCGACGTGATGCAGAACGAGGCCGCCCAGGTGCAGGCGATTACTTCCGCCATGC
TTACCTTTGGTATGGGCGCTTCCTCTATGGCTCTGTTTGCCCGTGTGGGCGGCGGCATCTTTACTAAGGCCGCC
GACGTAGGCGCCGACCTGGTTGGTAAGGTCGAGGCCGGCATCCCCGAGGACGACCCCCGCAACCCCGCCGTT
ATCGCCGATAACGTAGGCGACAACGTAGGCGACGTTGCCGGCATGGGCGCCGACCTGTACGAATCCTATGTGG
GCTCCATTATTTCCACCGCCGCTCTTCGGCCGTGGCCGCCGGCTTAGGCTTTAAGGGCGTGGTGATCCCCATGTC
CATGGCCGCTATTGGCATTCTGGCTTCCATCATCGGCACCTTCTTTGTACGCACCAAGGAAAACGCCTCTCAGA
AAAACCTTCTGGGATCTCTGCGGCTTGGAACCTGGATTTCCGCCGCTTTGATCGCTGTGGCCGCTTTCCCGCTG
ATTTACTTCGGCCTGGGCGCGGAAAAGATCGGCTTCTATTTCGCCATTCTTTCCGGCCTGCTGGCTGGTATTCT
CATCGGTGTGGTCACCGAATATTACACCTCCGATTCCTATAAGCCTACCAAGAAGCTAGCGGGCACCTCCAGCA
CCGGCCCCGCTACCGTCATCATCGGCGGCCTGTCCCTGGGAATGCTTTCCACCGTGCTTCCTGTAGTCATTGTC
GGCGCTTCCGTTTTGGTCAGCTATTTTCTGTCCGGCGGCAGCGACTCCTATAATAACGGCCTCTACGGCGTCGG
CCTGTCCGCCGTCGGTATGCTGGCTACCCTGGGCATTACCCTGGCCACCGACGCTTACGGCCCGGTCGCCGAT
AACGCCGGCGGTATCGCGGAGATGGCCCACATGGGCGAGGAGGTCCGCAAGCGCACCGACGCTTTGGATTCC
CTGGGCAACACCACTGCCGCTACCGGCAAGGGCTTTGCCATCGGCTCCGCCGCTCTCACCGCCCTGGCCTTGA
TCGCTTCCTATATCGCGGAAATCCAGGTGCTGAAGCCTGATTTCAACTTTAACCTGACCATTACCAATCCCCCCG
TACTCATCGGCCTGTTCATCGGCGGCGTGCTGCCGTTCCTGTTCGCCGCCCTCACCATGGATGCGGTCGGCCG
CTGCGCCCAGAAGATCGTGGTGGAGGTCCGCCGCCAGTTCCGTGAGATCAAGGGACTGATGACCGGGGAAGC
GGATCCTGATTACGCTTCCTGCGTGGATATCTGCACTCGTTCCGCCCAGAGGGAAATGATCGCTCCCGCTATCG
TGGCGGTAGCCGCCCCCATCATCGTGGGCCTGATCCTGGGCGTAAACGGCGTTGCCGGTATGCTGGCCGGCG
CCACAGTCTCCGGATTTATTCTCGCCGTTATGATGGCAAACTCCGGCGGCGCCTGGGATAACGCTAAGAAGTAC
ATTGAAAGCGGCGAGTATGGCGGAAAAGGCTCCGATAACCATAAGGCCGCCGTTGTGGGCGATACCGTGGGC
GATCCCTTCAAGGATACCTCCGGCCCGTCTATCAACATCCTGATCAAGCTGCTGAGCATGGTTTCCATCGTGTT
CGCCAGCGTGGTGCTTTCCTTCAGCCTGATTAAATAACTGCAGGAATTCGATATCAAGCTTATCGAT

>pDDGFP2-*Cl*-PPase

GACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGTATGATCCAAT
ATCAAAGGAAATGATAGCATTGAAGGATGAGACTAATCCAATTGAGGAGTGGCAGCATATAGAACAGCTAAAGG
GTAGTGCTGAAGGAAGCATACGATACCCCGCATGGAATGGGATAATATCACAGGAGGTACTAGACTACCTTTCA
TCCTACATAAATAGACGCATATAAGTACGCATTTAAGCATAAACACGCACTATGCCGTTCTTCTCATGTATATATA
TATACAGGCAACACGCAGATATAGGTGCGACGTGAACAGTGAGCTGTATGTGCGCAGCTCGCGTTGCATTTTCG
GAAGCGCTCGTTTTCGGAAACGCTTTGAAGTTCCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCA
GAGCGCTTTTGAAAACCAAAAGCGCTCTGAAGACGCACTTTCAAAAAACCAAAAACGCACCGGACTGTAACGAG
CTACTAAAATATTGCGAATACCGCTTCCACAAACATTGCTCAAAAGTATCTCTTTGCTATATATCTCTGTGCTATAT
CCCTATATAACCTACCCATCCACCTTTCGCTCCTTGAACTTGCATCTAAACTCGACCTCTACATTTTTTATGTTTAT
CTCTAGTATTACTCTTTAGACAAAAAAAATTGTAGTAAGAACTATTCATAGAGTGAATCGAAAACAATACGAAAATG
TAAACATTTCCTATACGTAGTATATAGAGACAAAATAGAAGAAACCGTTCATAATTTTCTGACCAATGAAGAATCA

TCAACGCTATCACTTTCTGTTCACAAAGTATGCGCAATCCACATCGGTATAGAATATAATCGGGGATGCCTTTAT
CTTGAAAAAATGCACCCGCAGCTTCGCTAGTAATCAGTAAACGCGGGAAGTGGAGTCAGGCTTTTTTTATGGAA
GAGAAAATAGACACCAAAGTAGCCTTCTTCTAACCTTAACGGACCTACAGTGCAAAAAGTTATCAAGAGACTGCA
TTATAGAGCGCACAAAGGAGAAAAAAAGTAATCTAAGATGCTTTGTTAGAAAAATAGCGCTCTCGGGATGCATTT
TTGTAGAACAAAAAAGAAGTATAGATTCTTTGTTGGTAAAATAGCGCTCTCGCGTTGCATTTCTGTTCTGTAAAAA
TGCAGCTCAGATTCTTTGTTTGAAAAATTAGCGCTCTCGCGTTGCATTTTTGTTTTACAAAAATGAAGCACAGATT
CTTCGTTGGTAAAATAGCGCTTTCGCGTTGCATTTCTGTTCTGTAAAAATGCAGCTCAGATTCTTTGTTTGAAAAA
TTAGCGCTCTCGCGTTGCATTTTTGTTCTACAAAATGAAGCACAGATGCTTCGTTCAGGTGGCACTTTTCGGGGA
AATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCT
GATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTT
TGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGG
GTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGT
TTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAA
CTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGAT
GGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGAC
AACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTT
GGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAACAAC
GTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGG
ATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGT
GAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACA
CGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCA
TTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTA
GGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCC
CGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCA
CCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAG
AGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGC
CTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTG
GACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGC
TTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAG
GGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGG
GGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCG
TCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTT
TTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATA
CCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGC
AAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGG
GCAGTGAGCGCAACGCAATTAATGTGAGTTACCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCG
GCTCCTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGCCAAG
CGCGCAATTAACCCTCACTAAAGGGAACAAAAGCTGGAGCTCACGGATTAGAAGCCGCCGAGCGGGTGACAGC
CCTCCGAAGGAAGACTCTCCTCCGTGCGTCCTCGTCTTCACCGGTCGCGTTCCTGAAACGCAGATGTGCCTCG
CGCCGCACTGCTCCGAACAATAAAGATTCTACAATACTAGCTTTTATGGTTATGAAGAGGAAAAATTGGCAGTAA
CCTGGCCCCACAAACCTTCAAATGAACGAATCAAATTAACAACCATAGGATGATAATGCGATTAGTTTTTTAGCCT
TATTTCTGGGGTAATTAATCAGCGAAGCGATGATTTTTGATCTATTAACAGATATATAAATGCAAAAACTGCATAA
CCACTTTAACTAATACTTTCAACATTTTCGGTTTGTATTACTTCTTATTCAAATGTAATAAAAGTATCAACAAAAAAT
TGTTAATATACCTCTATACTTTAACGTCAAGGAGAAAAAACCCCGGATTCTAGAACTAGTGGATCCCCCATGCAC
CATCACCATCACCATCACCATAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGAT
GGTGATGTTAATGGGCACAAATTTTCTGTCCGTGGAGAGGGTGAAGGTGATGCTACAAACGGAAAACTCACCCT
TAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACACTTGTCACTACTCTGACCTATGGTGTTCAA
TGCTTTTCCCGTTATCCGGATCACATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAG
GAACGCACTATATCTTTCAAAGATGACGGGACCTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCT
TGTTAATCGTATCGAGTTAAAAGGTATTGATTTTAAAGAAGATGGAAACATTCTCGGACACAAACTCGAGTACAA
CTTTAACTCACACAATGTATACATCACGGCAGACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCA
CAACGTTGAAGATGGTTCCGTTCAACTAGCAGACCATTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCT
TTTACCAGACAACCATTACCTGTCGACACAATCTGTCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGT
CCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTACAAGACCTCTGCTGGTGAGA
ACTTGTACTTTCAAGGTACTTCTGCTGGTGACCTGTCCATCCTTGCTCCTGTCGGCGCGGTAGTGGCTCTGCTG
TTCGCGTTTTTCCTTTCCCGCAGAGTTATGAAGGCTGACGAGGGCACCGACCTGATGAAGAAAATCGCCGGCG
CTGTCCGGAAAGGCGCCAACGCTTATCTGAAGCGCCAGTATACCGGCGTGGCTATTTTCTTCGCGGTTATGTTC
GTTATCCTGCTGATCCTGGCCATCGCCGGGTTCCTGACCCCCTTTGTGCCGGTTGCGTTCCTGACCGGAGGCT
TCTTCTCCGCCTTGTCCGGCTTTATCGGCATGAAGATCGCGACCTATGCCAACGGCCGCACCGCCAACAGCGC
CAGCAAAAGCCTGAACGCGGGCCTGAGAATCGCTTTCTCCGCCGGCGGCGTAATGGGCTTTGTGGTGGTAGGC
CTGGGCCTGCTGGATATCTCTATCTGGTATTTCTTCCTGAAGTTCTGGTACAGCACTGTCGACGTGATGCAGAAC
GAGGCCGCCCAGGTGCAGGCGATTACTTCCGCCATGCTTACCTTTGGTATGGGCGCTTCCTCTATGGCTCTGTT
TGCCCGTGTGGGCGGCGGCATCTTTACTAAGGCCGCCGACGTAGGCGCCGACCTGGTTGGTAAGGTCGAGGC
CGGCATCCCCGAGGACGACCCCCGCAACCCCGCCGTTATCGCCGATAACGTAGGCGACAACGTAGGCGACGT
TGCCGGCATGGGCGCCGACCTGTACGAATCCTATGTGGGCTCCATTATTTCCACCGCCGCTCTGGCCGTGGCC
GCCGGCTTAGGCTTTAAGGGCGTGGTGATCCCCATGTCCATGGCCGCTATTGGCATTCTGGCTTCCATCATCGG
CACCTTCTTTGTACGCACCAAGGAAAACGCCTCTCAGAAAAACCTTCTGGGATCTCTGCGGCTTGGAACCTGGA
TTTCCGCCGCTTTGATCGCTGTGGCCGCTTTCCCGCTGATTTACTTCGGCCTGGGCGCGGAAAAGATCGGCTTC
TATTTCGCCATTCTTTCCGGCCTGCTGGCTGGTATTCTCATCGGTGTGGTCACCGAATATTACACCTCCGATTCC
TATAAGCCTACCAAGAAGCTAGCGGGCACCTCCAGCACCGGCCCCGCTACCGTCATCATCGGCGGCCTGTCCC
TGGGAATGCTTTCCACCGTGCTTCCTGTAGTCATTGTCGGCGCTTCCGTTTTGGTCAGCTATTTTCTGTCCGGCG
GCAGCGACTCCTATAATAACGGCCTCTACGGCGTCGGCCTGTCCGCCGTCGGTATGCTGGCTACCCTGGGCAT
TACCCTGGCCACCGACGCTTACGGCCCGGTCGCCGATAACGCCGGCGGTATCGCGGAGATGGCCCACATGGG
CGAGGAGGTCCGCAAGCGCACCGACGCTTTGGATTCCCTGGGCAACACCACTGCCGCTACCGGCAAGGGCTT
TGCCATCGGCTCCGCCGCTCTCACCGCCCTGGCCTTGATCGCTTCCTATATCGCGGAAATCCAGGTGCTGAAG
CCTGATTTCAACTTTAACCTGACCATTACCAATCCCCCCGTACTCATCGGCCTGTTCATCGGCGGCGTGCTGCC
GTTCCTGTTCGCCGCCCTCACCATGGATGCGGTCGGCCGCTGCGCCCAGAAGATCGTGGTGGAGGTCCGCCG
CCAGTTCCGTGAGATCAAGGGACTGATGACCGGGGAAGCGGATCCTGATTACGCTTCCTGCGTGGATATCTGC

ACTCGTTCCGCCCAGAGGGAAATGATCGCTCCCGCTATCGTGGCGGTAGCCGCCCCCATCATCGTGGGCCTGA
TCCTGGGCGTAAACGGCGTTGCCGGTATGCTGGCCGGCGCCACAGTCTCCGGATTTATTCTCGCCGTTATGAT
GGCAAACTCCGGCGGCGCCTGGGATAACGCTAAGAAGTACATTGAAAGCGGCGAGTATGGCGGAAAAGGCTC
CGATAACCATAAGGCCGCCGTTGTGGGCGATACCGTGGGCGATCCCTTCAAGGATACCTCCGGCCCGTCTATC
AACATCCTGATCAAGCTGCTGAGCATGGTTTCCATCGTGTTCGCCAGCGTGGTGCTTTCCTTCAGCCTGATTAAA
TAACTGCAGGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAGTCATGTAATTAGTTATGTCACGCTTAC
ATTCACGCCCCTCCCCCCACATCCGCTCTAACCGAAAAGGGAAGGAGTTAGACAACCTGAAGTCTAGGTCCCCT
ATTTATTTTTTTATAGTTATGTTAGTATTAAGAACGTTATTTATATTCAAATTTTCCTTTTTTTTCTGTACAGACGCGT
GTACGCATGTAACATTATACTGAAAACCTTGCTTGAGAAGGTTTTGGGACCCAATTCGCCCTATAGTGAGTCGTA
TTACGCGCGCTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCC
TTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTT
GCGCAGCCTGAATGGCGAATGGCGCGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTAC
GCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCA
CGTTCGCCGGCAAATGGTCAGGTCATTGAGTGTTTTTTATTTGTTGTATTTTTTTTTTTTTAGAGAAAATCCTCCAA
TATCAAATTAGGAATCGTAGTTTCATGATTTTCTGTTACACCTAACTTTTTGTGTGGTGCCCTCCTCCTTGTCAATA
TTAATGTTAAAGTGCAATTCTTTTTCCTTATCACGTTGAGCCATTAGTATCAATTTGCTTACCTGTATTCCTTTACT
ATCCTCCTTTTTCTCCTTCTTGATAAATGTATGTAGATTGCGTATATAGTTTCGTCTACCCTATGAACATATTCCAT
TTTGTAATTTCGTGTCGTTTCTATTATGAATTTCATTTATAAAGTTTATGTACAAATATCATAAAAAAAGAGAATCTT
TTTAAGCAAGGATTTTCTTAACTTCTTCGGCGACAGCATCACCGACTTCGGTGGTACTGTTGGAACCACCTAAAT
CACCAGTTCTGATACCTGCATCCAAAACCTTTTTAACTGCATCTTCAATGGCCTTACCTTCTTCAGGCAAGTTCAA
TGACAATTTCAACATCATTGCAGCAGACAAGATAGTGGCGATAGGGTCAACCTTATTCTTTGGCAAATCTGGAGC
AGAACCGTGGCATGGTTCGTACAAACCAAATGCGGTGTTCTTGTCTGGCAAAGAGGCCAAGGACGCAGATGGC
AACAAACCCAAGGAACCTGGGATAACGGAGGCTTCATCGGAGATGATATCACCAAACATGTTGCTGGTGATTAT
AATACCATTTAGGTGGGTTGGGTTCTTAACTAGGATCATGGCGGCAGAATCAATCAATTGATGTTGAACCTTCAA
TGTAGGGAATTCGTTCTTGATGGTTTCCTCCACAGTTTTTCTCCATAATCTTGAAGAGGCCAAAAGATTAGCTTTA
TCCAAGGACCAAATAGGCAATGGTGGCTCATGTTGTAGGGCCATGAAAGCGGCCATTCTTGTGATTCTTTGCAC
TTCTGGAACGGTGTATTGTTCACTATCCCAAGCGACACCATCACCATCGTCTTCCTTTCTCTTACCAAAGTAAATA
CCTCCCACTAATTCTCTGACAACAACGAAGTCAGTACCTTTAGCAAATTGTGGCTTGATTGGAGATAAGTCTAAA
AGAGAGTCGGATGCAAAGTTACATGGTCTTAAGTTGGCGTACAATTGAAGTTCTTTACGGATTTTTAGTAAACCTT
GTTCAGGTCTAACACTACCGGTACCCCATTTAGGACCAGCCACAGCACCTAACAAAACGGCATCAACCTTCTTG
GAGGCTTCCAGCGCCTCATCTGGAAGTGGGACACCTGTAGCATCGATAGCAGCACCACCAATTAAATGATTTTC
GAAATCGAACTTGACATTGGAACGAACATCAGAAATAGCTTTAAGAACCTTAATGGCTTCGGCTGTGATTTCTTG
ACCAACGTGGTCACCTGGCAAAACGACGATCTTCTTAGGGGCAGACATTACAATGGTATATCCTTGAAATATATA
TAAAAAAAGGCGCCTTAGACCGCTCGGCCAAACAACGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCC
CTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTG
GGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCC
AAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTG
GTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCCTGATGCG
GTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATAGGGTAATAACTGATATAATTAAATTGAAGCTCT
AATTTGTGAGTTTAGTATACATGCATTTACTTATAATACAGTTTTTTAGTTTTGCTGGCCGCATCTTCTCAAATATG
CTTCCCAGCCTGCTTTTCTGTAACGTTCACCCTCTACCTTAGCATCCCTTCCCTTTGCAAATAGTCCTCTTCCAAC
AATAATAATGTCAGATCCTGTAGAGACCACATCATCCACGGTTCTATACTGTTGACCCAATGCGTCTCCCTTGTC
ATCTAAACCCACACCGGGTGTCATAATCAACCAATCGTAACCTTCATCTCTTCCACCCATGTCTCTTTGAGCAATA
AAGCCGATAACAAAATCTTTGTCGCTCTTCGCAATGTCAACAGTACCCTTAGTATATTCTCCAGTAGATAGGGAG
CCCTTGCATGACAATTCTGCTAACATCAAAAGGCCTCTAGGTTCCTTTGTTACTTCTTCTGCCGCCTGCTTCAAA
CCGCTAACAATACCTGGGCCCACCACACCGTGTGCATTCGTAATGTCTGCCCATTCTGCTATTCTGTATACACCC
GCAGAGTACTGCAATTTGACTGTATTACCAATGTCAGCAAATTTTCTGTCTTCGAAGAGTAAAAAATTGTACTTGG
CGGATAATGCCTTTAGCGGCTTAACTGTGCCCTCCATGGAAAAATCAGTCAAGATATCCACATGTGTTTTTAGTA
AACAAATTTTGGGACCTAATGCTTCAACTAACTCCAGTAATTCCTTGGTGGTACGAACATCCAATGAAGCACACA
AGTTTGTTTGCTTTTCGTGCATGATATTAAATAGCTTGGCAGCAACAGGACTAGGATGAGTAGCAGCACGTTCCT
TATATGTAGCTTTCGACATGATTTATCTTCGTTTCCTGCAGGTTTTTGTTCTGTGCAGTTGGGTTAAGAATACTGG
GCAATTTCATGTTTCTTCAACACTACATATGCGTATATATACCAATCTAAGTCTGTGCTCCTTCCTTCGTTCTTCCT
TCTGTTCGGAGATTACCGAATCAAAAAAATTTCAAAGAAACCGAAATCAAAAAAAAGAATAAAAAAAAAATGATGA
ATTGAATTGAAAAGCTGTGGTATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCC
GACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGT
GACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGA

>pZH-*Cl*-PPase

ATGAAAATAAAAACAGGTGCACGCATCCTCGCATTATCCGCATTAACGACGATGATGTTTTCCGCCTCGGCTCTC
GCCCGCGGCTCCCATCATCACCATCACCATCACCATGAAAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAA
CGGCGATAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCG
TTGAGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTC
TGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTTCC
AGGACAAGCTGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCGCTGTT
GAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATCCCGGCGCT
GGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTTCACCTGGCCG
CTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAGACGTGGGCGTGG
ATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACC
GATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTC
CAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCAAGGGTCAACCATCCAAACCGT
TCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAGCTGGCAAAAGAGTTCCTCGAAAA
CTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTT
ACGAGGAAGAGTTGGTGAAAGATCCGCGGATTGCCGCCACTATGGAAAACGCCCAGAAAGGTGAAATCATGCC
GAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAG
ACTGTCGATGAAGCCCTGAAAGACGCGCAGACCCATGGTAATGCGAGCTCGAACAACAACAACAATAACAATAA
CAACAACGCGAGCTCGCTGGAAGTTCTGTTCCAGGGTCCGACTAGTGGCAGCGGCAGCGCTAGCCTGTCCATC
CTTGCTCCTGTCGGCGCGGTAGTGGCTCTGCTGTTCGCGTTTTTCCTTTCCCGCAGAGTTATGAAGGCTGACGA
GGGCACCGACCTGATGAAGAAAATCGCCGGCGCTGTCCGGAAAGGCGCCAACGCTTATCTGAAGCGCCAGTAT
ACCGGCGTGGCTATTTTCTTCGCGGTTATGTTCGTTATCCTGCTGATCCTGGCCATCGCCGGGTTCCTGACCCC

CTTTGTGCCGGTTGCGTTCCTGACCGGAGGCTTCTTCTCCGCCTTGTCCGGCTTTATCGGCATGAAGATCGCGA
CCTATGCCAACGGCCGCACCGCCAACAGCGCCAGCAAAAGCCTGAACGCGGGCCTGAGAATCGCTTTCTCCGC
CGGCGGCGTAATGGGCTTTGTGGTGGTAGGCCTGGGCCTGCTGGATATCTCTATCTGGTATTTCTTCCTGAAGT
TCTGGTACAGCACTGTCGACGTGATGCAGAACGAGGCCGCCCAGGTGCAGGCGATTACTTCCGCCATGCTTAC
CTTTGGTATGGGCGCTTCCTCTATGGCTCTGTTTGCCCGTGTGGGCGGCGGCATCTTTACTAAGGCCGCCGAC
GTAGGCGCCGACCTGGTTGGTAAGGTCGAGGCCGGCATCCCCGAGGACGACCCCCGCAACCCCGCCGTTATC
GCCGATAACGTAGGCGACAACGTAGGCGACGTTGCCGGCATGGGCGCCGACCTGTACGAATCCTATGTGGGC
TCCATTATTTCCACCGCCGCTCTGGCCGTGGCCGCCGGCTTAGGCTTTAAGGGCGTGGTGATCCCCATGTCCA
TGGCCGCTATTGGCATTCTGGCTTCCATCATCGGCACCTTCTTTGTACGCACCAAGGAAAACGCCTCTCAGAAA
AACCTTCTGGGATCTCTGCGGCTTGGAACCTGGATTTCCGCCGCTTTGATCGCTGTGGCCGCTTTCCCGCTGAT
TTACTTCGGCCTGGGCGCGGAAAAGATCGGCTTCTATTTCGCCATTCTTTCCGGCCTGCTGGCTGGTATTCTCA
TCGGTGTGGTCACCGAATATTACACCTCCGATTCCTATAAGCCTACCAAGAAGCTAGCGGGCACCTCCAGCACC
GGCCCCGCTACCGTCATCATCGGCGGCCTGTCCCTGGGAATGCTTTCCACCGTGCTTCCTGTAGTCATTGTCG
GCGCTTCCGTTTTGGTCAGCTATTTTCTGTCCGGCGGCAGCGACTCCTATAATAACGGCCTCTACGGCGTCGGC
CTGTCCGCCGTCGGTATGCTGGCTACCCTGGGCATTACCCTGGCCACCGACGCTTACGGCCCGGTCGCCGATA
ACGCCGGCGGTATCGCGGAGATGGCCCACATGGGCGAGGAGGTCCGCAAGCGCACCGACGCTTTGGATTCCC
TGGGCAACACCACTGCCGCTACCGGCAAGGGCTTTGCCATCGGCTCCGCCGCTCTCACCGCCCTGGCCTTGAT
CGCTTCCTATATCGCGGAAATCCAGGTGCTGAAGCCTGATTTCAACTTTAACCTGACCATTACCAATCCCCCCGT
ACTCATCGGCCTGTTCATCGGCGGCGTGCTGCCGTTCCTGTTCGCCGCCCTCACCATGGATGCGGTCGGCCGC
TGCGCCCAGAAGATCGTGGTGGAGGTCCGCCGCCAGTTCCGTGAGATCAAGGGACTGATGACCGGGGAAGCG
GATCCTGATTACGCTTCCTGCGTGGATATCTGCACTCGTTCCGCCCAGAGGGAAATGATCGCTCCCGCTATCGT
GGCGGTAGCCGCCCCCATCATCGTGGGCCTGATCCTGGGCGTAAACGGCGTTGCCGGTATGCTGGCCGGCGC
CACAGTCTCCGGATTTATTCTCGCCGTTATGATGGCAAACTCCGGCGGCGCCTGGGATAACGCTAAGAAGTACA
TTGAAAGCGGCGAGTATGGCGGAAAAGGCTCCGATAACCATAAGGCCGCCGTTGTGGGCGATACCGTGGGCG
ATCCCTTCAAGGATACCTCCGGCCCGTCTATCAACATCCTGATCAAGCTGCTGAGCATGGTTTCCATCGTGTTC
GCCAGCGTGGTGCTTTCCTTCAGCCTGATTAAACCTGCAGGCTAATGCAAGCTTGCGGCCGCACTCGAGCACC
ACCACCACCACCACTGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGA
GCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATAT
CCGGATTGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGT
GACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCG
GCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGACCCC
AAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTT
GGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTT
GATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATT
TTAACAAAATATTAACGTTTACAATTTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATT
TTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCTTAGAAAAACTCATCGAGCATCAAATGAAACTGC
AATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGA
GGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATT
AATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGG
CAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCA
ACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAA
ACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTC
TTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAA
AATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGG
CAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCAC
CTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCC
TAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTAT
TGTTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGAT
CTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTT
GTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACT
GTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCT
AATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTAC
CGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACA
CCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTA
TCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTA
TAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTAT
GGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTG
CGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACG
ACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGT
GCGGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATAC
ACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGA
CGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGT
TTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACA
GATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGC
GGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGG
GTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGA
ACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAG
CGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAAT
GGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTC
AGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTA
AGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATG
CCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCG
TGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAA
TGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGAT
AGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGG
TGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGT
GCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTC
TTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTC

CACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTT
CGGTATCGTCGTATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGC
GCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTT
GTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATAT
TTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGT
GACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGT
GTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTC
ATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTT
CGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGC
GACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGC
CAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCG
CAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATC
GTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAA
GGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCC
AGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGT
CCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGA
TCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACG
ATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCG
GATAACAATTCCCCTCTAGAAATAATTTTATTTAACTTTAAGAAAGGAAACAGCG

>pMP649-*Pa*-PPase

ATACCTTTTAATGAAAAGGGAGAAAAGGAAAGAGCCGGAACCGGCTTTCATGAAGAAGCTCATGACTAATGCTT
GCATCACAATACTGAAGTTGACATATTTAAGGTGTTTTTAATAGGTGGTTAGCAATCGTTTACTTTCTAACTTTTCT
TCCTTTTACATTTCAGCAATATATATATATATTTCAAGGATATACCATTCTAATGTCTGCCCCTAAGAAGATCGTCG
TTTTGCCAGGTGACCACGTTGGTCAAGAAATCACAGCCGAAGCCATTAAGGTTCTTAAAGCTATTTCTGATGTTC
GTTCCAATGTCAAGTTCGATTTCGAAAATCATTTAATTGGTGGTGCTGCTATCGATGCTACAGGTGTTCCACTTC
CAGATGAGGCGCTGGAAGCCTCCAAGAAGGCTGATGCCGTTTTGTTAGGTGCTGTGGGTGGTCCTAAATGGGG
TACCGGTAGTGTTAGACCTGAACAAGGTTTACTAAAAATCCGTAAAGAACTTCAATTGTACGCCAACTTAAGACC
ATGTAACTTTGCATCCGACTCTCTTTTAGACTTATCTCCAATCAAGCCACAATTTGCTAAAGGTACTGACTTCGTT
GTTGTCAGAGAATTAGTGGGAGGTATTTACTTTGGTAAGAGAAAGGAAGACGATGGTGATGGTGTCGCTTGGGA
TAGTGAACAATACACCGTTCCAGAAGTGCAAAGAATCACAAGAATGGCCGCTTTCATGGCCCTACAACATGAGC
CACCATTGCCTATTTGGTCCTTGGATAAAGCTAATGTTTTGGCCTCTTCAAGATTATGGAGAAAAACTGTGGAGG
AAACCATCAAGAACGAATTCCCTACATTGAAGGTTCAACATCAATTGATTGATTCTGCCGCCATGATCCTAGTTAA
GAACCCAACCCACCTAAATGGTATTATAATCACCAGCAACATGTTTGGTGATATCATCTCCGATGAAGCCTCCGT
TATCCCAGGTTCCTTGGGTTTGTTGCCATCTGCGTCCTTGGCCTCTTTGCCAGACAAGAACACCGCATTTGGTTT
GTACGAACCATGCCACGGTTCTGCTCCAGATTTGCCAAAGAATAAGGTCAACCCTATCGCCACTATCTTGTCTGC
TGCAATGATGTTGAAATTGTCATTGAACTTGCCTGAAGAAGGTAAGGCCATTGAAGATGCAGTTAAAAAGGTTTT
GGATGCAGGTATCAGAACTGGTGATTTAGGTGGTTCCAACAGTACCACCGAAGTCGGTGATGCTGTCGCCGAA
GAAGTTAAGAAAATCCTTGCTTAAAAAGATTCTCTTTTTTTATGATATTTGTACATAAACTTTATAAATGAAATTCAT
AATAGAAACGACACGAAATTACAAAATGGAATATGTTCATAGTAGACGAAACTATATACGCAATCTACATACATTT
ATCAAGAAGGAGAAAAAGGAGGATGTAAAGGAATACAGGTAAGCAAATTGATACTAATGGCTCAACGTGATAAG
GAAAAAGAATTGCACTTTAACATTAATATTGACAAGGAGGAGGGCACCACACAAAAAGTTAGGTGTAACAGAAAA
TCATGAAACTATGATTCCTAATTTATATATTGGAGGATTTTCTCTAAAAAAAAAAAAAATACAACAAATAAAAAACAC
TCAATGACCTGACCATTTGATGGAGTTTAAGTCAATACCTTCTTGAACCATTTCCCATAATGGTGAAAGTTCCCTC
AAGAATTTTACTCTGTCAGAAACGGCCTTAACGACGTACCGGATCTATGCGGTGTGAAATACCGCACAGATGCG
TAAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCG
GGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGG
TTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGCCAAGCTAGCTTCCTGAAACGGAGAAACATAAACAGG
CATTGCTGGGATCACCCATACATCACTCTGTTTTGCCTGACCTTTTCCGGTAATTTGAAAACAAACCCGGTCTCG
AAGCGGAGATCCGGCGATAATTACCGCAGAAATAAACCCATACACGAGACGTAGAACCAGCCGCACATGGCCG
GAGAAACTCCTGCGAGAATTTCGTAAACTCGCGCGCATTGCATCTGTATTTCCTAATGCGGCACTTCAGGCCTC
GATCGAGACCGTTTATCCATTGCTTTTTTGTTGTCTTTTTCCCTCGTTCACAGAAAGTCTGAAGAAGCTATAGTAG
AACTATGAGCTTTTTTTGTTTCTGTTTTCCTTTTTTTTTTTTTTACCTCTGTGGAAATTGTTACTCTCACACTCTTTA
GTTCGTTTGTTTGTTTTGTTTATTCCAATTATGACCGGTGACGAAACGTGGTCGATGGTGGGTACCGCTTATGCT
CCCCTCCATTAGTTTCGATTATATAAAAAGGCCAAATATTGTATTATTTTCAAATGTCCTATCATTATCGTCTAACA
TTAATTTCTCTTAAATTTTTTCTCTTTCTTTCCTATAACACCAATAGTGAAAATCTTTTTTTTCTTCTATATCTACAAAA
ACTTTTTTTTTTCTATCAACCTCGTTGATAAATTTTTTCTTTAACAATCGTTAATAATTAATTAATTGGAAAATAACCA
TTTTTTCTCTCTTTTATACACACATTCAAAAGAAAGAAAAAAAATATACCCCAGCCTCGACATGCATCACCATCAC
CATCACGGTGGAAATATGATAAGCTATGCCTTACTAGGCGTCATTTTAGGAATTTCCGGAGTTATATATGCGGTG
TATTTGGCTGTGTGGGTTTTGAGACAAGACCCTGGTAACGAGAAAATGCGCTTTATTTCACAGGCCATAGCCACT
GGCGCGAGGGCTTATCTCTTCAGGCAGTACCGGACGCTGGCAGTTCTACTAGTTATACTCGCCGTATTAATATT
AGTTGCAATTGATATGCCGCGCAGAACGTTTGGGCTAACCGCTCTGGCTTTTATCGTGGGGGCGCTGGGCTCC
ATGCTCGCCGGCTATTTGGGCATGTACGTGACGACGAGATCAGCCTCTCGCGTCGCCCAGGCGGCGGCCACG
GGGGGAATGGGCAAGGCGTTATTAGTGTCGTGGCGCGCGGGGGCTGTCATGGGCCTCTCCCTGGCCAGCATT
GCGTTGCTTCTTATCTCGGGGTTTTACCTCGTGTTCAGATCTGTACTTCCCGATGATTGGGCCGTGCCTTTAGTG
GCCTTAGGCTTTGGAGCCTCGCTGGTTACTTTGTTTATGAGAGTTGGCGGCGGCATATATACAAAGGCGGCGGA
TCTCGGCGCAGATTTAGTTGGGAAGGTAGAGGCCGGCATTCCTGAGGACGACCCCCGCAACCCAGGGGTAATT
GCGGATAACGTCGGCGACAACGTTGGAGATGTGGCCGGCATGGCGGCAGACGTCTACGAGTCGTATATAGTAA
CAGTAACTGCGGCCATATTCCTTGCCGCCATCCTCGGTCTGCCCACGCAATTTATTGAGGCAATAATCCTCTTCG
CCGCGTTGGCGCTTGTGGCGACTTTTGCGGGAGTAAATCTCCTGAAGACTACAGGTGTTAAACACCCGCTTTCG
TCCATTAGTTTAGCTATATATGCAACAATTGGGCTTTCTGTTGTGCTGTTTTTCATAGGGGCATTCACGTTAGGAC
TGGACTCCACGAAAGCGCTGGCCTTGGCGGCAACTACGTCTCTTGGCGCAGTGATCGCACCGCTGATTGTAAA
AATAACAGATTACTACACGTCCTATAACTACGGCCCAGTTAGAAAAATAGCAGAGCAGGCAAAGATCAGCCCGG
CCACCGTAATAATTACGGGTTATGGCGTGGGATTAATGAGCGCAATACCTGTGATAGCGGTTATTGTCGCCGTA
CTGGGCATATCCTACATGATAGGTTATTACACTGTGCCTGTAAGCGGCTTTGGCGAGCTTTCTAAATATCTCGCT
GGGATATTCGGCACGGCCATGGCCAGCGTGGGACTGCTCGTAGTGGCGGGGATTATAATAACTGCAGACTCCT
ACGGCCCTGTAAGCGATAATGCTGGCGGCGTTGTGGAAATGGCGGGGTTGCCAGATGAAGTGAGGGAGATCA
CAGATGTACTGGACTCCGTCGGCAATACCACAAAAGCCACTACTAAGGGATATGCAATAGCCAGCGCGGCGCT
CGCGGCGTTAGTCCTCTTCATCGCTTTAATTTTTGAAATAGTTTACTCGGCGTCGAAAATCTTAGGTAAAGGTATA

GTCGATATGATAAGCGAGAGCCTCTCAGGGCTTCAGCTGATTAACGCCAACGTGTTAATAGGCGCCTTCTTAGG
AGTAGCGCTTGTGTACTTTTTCAGTAGCCGCACTCTCGAGGCCGTCGGCAGGACTGCCATGGAGATAGTTGAG
GAGATCAGAAGGCAGTTCCGCGAGAAGCCAGGCATCTTAGAGTGGAAAGAACAGCCGGATTACGCCCGCGTTG
TAGACATCGCCACCAGAAGGGCGTTGGGAGAGTTCTTAATCCCCGGCCTTGCGGCCATTGTCCTCCCCTTAATA
ACCGGCTTGTTGCTTGGGTGGAATGCCCTGGCTGGCCTCATCATGGGGGCAATAGTAGCGGGGGTGCCCAGG
GCTTTGCTCATGGCTAACGCAGGCGGGGCGTGGGATAACGCCAAGAAGTATATAGAAATACAGGGATTGAAGA
AAACGGAGATGCATAAAGCTGCCGTTATTGGCGACACTGTAGGAGATCCCATGAAAGATACTGTTGGACCTTCT
CTAAACCCGCTTATCAAGGTGCTTAACACGCTTTCTGTGGTATTCACCTATGTGATCGTGTCTACGAATATAGCG
CTAGGAATTTGGCCGTCAGGTCTATTGCCTTTCTGATCTCTAGTTCTAGTTCTAGTCGAGGGAATTGTCATTCTAA
TATTTTATCCACACACACACCTTAAAATTTTTAGATTAAATGGCATCAACTCTTAGCTTCACACACACACACACAC
CGAAGCTGGTTGTTTTATTTGATTTGATATAATTGGTTTCTCTGGATGGTACTTTTTCTTTCTTGGTTATTTCCTAT
TTTAAAATATGAAACGCACACAAGTCATAATTATTCTAATAGAGCACAATTCACAACACGCACATTTCAACTTTAAT
ATTTTTTTAGAAACACTTTATTTAGTCTAATTCTTAATTTTTAATATATATAATGCACACACACTAATTTATTCATTAA
TTTTTTATTGAGTAGGATTTGAAAATATTTGGTATCTTTGCAAGATGTTTGTATAGAGGGACAAAGAATCGTCTTTA
TTATGGTCAAGGCTTTACGTCATAATAGTTCCTGCCCAGCTCTTCTATAATACTTTAAAGATCTCTTCTCGTTTGC
TCCATTTGGAAGTCTCGCTTACGTTTATGCGCCCATACAGACACTCAAGATACACACTTACATGAACGTATACAA
ATTTACTAACACTACTTGAAAATATGAACCACAGTACATCATATTAAGACGTAGTATTCGATGATTGAAGGCCGCC
TCCGCGAAATACCTTTACTGATTTTGCCGGTTAATCGCATCGAAATTTCTTCATCACAAGAAAGCAAACAAATCGC
CAGGCCATTCTACAAGTTTCCTTTTCTTATGAAGATGTAAAAGCTACTAAGGCGTCATTACTCTACGAATTCGTAA
TCATGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAA
AGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAG
TCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGC
GCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTC
AAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAA
AAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACA
AAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAG
CTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCG
TGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTG
CACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGAC
ACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGA
GTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAG
TTACCTTCGGAAAAARAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTT
GCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCT
CAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTA
AATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAG
TGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTAC
GATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGAT
TTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCC
AGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTG
CTACAGGCATCGTSGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGA
GTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTG
GCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTT
TCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGC
GTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGC
GAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAG
CATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGG
GCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCAT
GAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCC
ACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTC
GCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAG
CGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAAC
TATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATACTAGAAAGTATAGGAACTTCAGAGCGCTTTTGAA
AACCAAAAGCGCTCTGAAGACGCACTTTCAAAAAACCAAAAACGCACCGGACTGTAACGAGCTACTAAAATATTG
CGAATACCGCTTCCACAAACATTGCTCAAAAGTATCTCTTTGCTATATATCTCTGTGCTATATCCCTATATAACCT
ACCCATCCACCTTTCGCTCCTTGAACTTGCATCTAAACTCGACCTACATTTTTTGTTTATCTAGTATTACTCTTTAG
ACAAAAAAATTGTAGTAATCATAGGTACAAAACGAAAGAATTTCTACGTATTTGGAATAGAAACTCT

>pMP649-*Tm*-PPase[†]

ATACCTTTTAATGAAAAGGGAGAAAAGGAAAGAGCCGGAACCGGCTTTCATGAAGAAGCTCATGACTAATGCTT
GCATCACAATACTGAAGTTGACATATTTAAGGTGTTTTTAATAGGTGGTTAGCAATCGTTTACTTTCTAACTTTTCT
TCCTTTTACATTTCAGCAATATATATATATATTTCAAGGATATACCATTCTAATGTCTGCCCCTAAGAAGATCGTCG
TTTTGCCAGGTGACCACGTTGGTCAAGAAATCACAGCCGAAGCCATTAAGGTTCTTAAAGCTATTTCTGATGTTC
GTTCCAATGTCAAGTTCGATTTCGAAAATCATTTAATTGGTGGTGCTGCTATCGATGCTACAGGTGTTCCACTTC
CAGATGAGGCGCTGGAAGCCTCCAAGAAGGCTGATGCCGTTTTGTTAGGTGCTGTGGGTGGTCCTAAATGGGG
TACCGGTAGTGTTAGACCTGAACAAGGTTTACTAAAAATCCGTAAAGAACTTCAATTGTACGCCAACTTAAGACC
ATGTAACTTTGCATCCGACTCTCTTTTAGACTTATCTCCAATCAAGCCACAATTTGCTAAAGGTACTGACTTCGTT
GTTGTCAGAGAATTAGTGGGAGGTATTTACTTTGGTAAGAGAAAGGAAGACGATGGTGATGGTGTCGCTTGGGA
TAGTGAACAATACACCGTTCCAGAAGTGCAAAGAATCACAAGAATGGCCGCTTTCATGGCCCTACAACATGAGC
CACCATTGCCTATTTGGTCCTTGGATAAAGCTAATGTTTTGGCCTCTTCAAGATTATGGAGAAAAACTGTGGAGG
AAACCATCAAGAACGAATTCCCTACATTGAAGGTTCAACATCAATTGATTGATTCTGCCGCCATGATCCTAGTTAA
GAACCCAACCCACCTAAATGGTATTATAATCACCAGCAACATGTTTGGTGATATCATCTCCGATGAAGCCTCCGT
TATCCCAGGTTCCTTGGGTTTGTTGCCATCTGCGTCCTTGGCCTCTTTGCCAGACAAGAACACCGCATTTGGTTT
GTACGAACCATGCCACGGTTCTGCTCCAGATTTGCCAAAGAATAAGGTCAACCCTATCGCCACTATCTTGTCTGC
TGCAATGATGTTGAAATTGTCATTGAACTTGCCTGAAGAAGGTAAGGCCATTGAAGATGCAGTTAAAAAGGTTTT
GGATGCAGGTATCAGAACTGGTGATTTAGGTGGTTCCAACAGTACCACCGAAGTCGGTGATGCTGTCGCCGAA
GAAGTTAAGAAAATCCTTGCTTAAAAAGATTCTCTTTTTTTATGATATTTGTACATAAACTTTATAAATGAAATTCAT
AATAGAAACGACACGAAATTACAAAATGGAATATGTTCATAGTAGACGAAACTATATACGCAATCTACATACATTT
ATCAAGAAGGAGAAAAAGGAGGATGTAAAGGAATACAGGTAAGCAAATTGATACTAATGGCTCAACGTGATAAG
GAAAAAGAATTGCACTTTAACATTAATATTGACAAGGAGGAGGGCACCACACAAAAAGTTAGGTGTAACAGAAAA

TCATGAAACTATGATTCCTAATTTATATATTGGAGGATTTTCTCTAAAAAAAAAAAAATACAACAAATAAAAAACAC
TCAATGACCTGACCATTTGATGGAGTTTAAGTCAATACCTTCTTGAACCATTTCCCATAATGGTGAAAGTTCCCTC
AAGAATTTTACTCTGTCAGAAACGGCCTTAACGACGTACCGGATCTATGCGGTGTGAAATACCGCACAGATGCG
TAAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCG
GGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGG
TTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGCCAAGCTAGCTTCCTGAAACGGAGAAACATAAACAGG
CATTGCTGGGATCACCCATACATCACTCTGTTTTGCCTGACCTTTTCCGGTAATTTGAAAACAAACCCGGTCTCG
AAGCGGAGATCCGGCGATAATTACCGCAGAAATAAACCCATACACGAGACGTAGAACCAGCCGCACATGGCCG
GAGAAACTCCTGCGAGAATTTCGTAAACTCGCGCGCATTGCATCTGTATTTCCTAATGCGGCACTTCAGGCCTC
GATCGAGACCGTTTATCCATTGCTTTTTTGTTGTCTTTTTCCCTCGTTCACAGAAAGTCTGAAGAAGCTATAGTAG
AACTATGAGCTTTTTTTGTTTCTGTTTTCCTTTTTTTTTTTTTTTACCTCTGTGGAAATTGTTACTCTCACACTCTTTA
GTTCGTTTGTTTGTTTTGTTTATTCCAATTATGACCGGTGACGAAACGTGGTCGATGGTGGGTACCGCTTATGCT
CCCCTCCATTAGTTTCGATTATATAAAAAGGCCAAATATTGTATTATTTTCAAATGTCCTATCATTATCGTCTAACA
TTAATTTCTCTTAAATTTTTTCTCTTTCTTTCCTATAACACCAATAGTGAAAATCTTTTTTTCTTCTATATCTACAAAA
ACTTTTTTTTTCTATCAACCTCGTTGATAAATTTTTTCTTTAACAATCGTTAATAATTAATTAATTGGAAAATAACCA
TTTTTTCTCTCTTTTATACACACATTCAAAAGAAAGAAAAAAAAATATACCCCAGCCTCGACATGCGAGGATCACAT
CATCATCATCATCATTACGTCGCTGCTCTTTTCTTTTTAATTCCGCTTGTTGCACTCGGTTTTGCTGCCGCAAACT
TCGCAGCCGTCGTCAGAAAACCAGAGGGGACCGAGCGGATGAAAGAGATTTCCTCCTACATCAGAAGTGGAGC
AGATTCCTTCCTTGCACACGAAACAAAAGCTATTTTCAAAGTTGCCATCGTCATAGCGATACTCCTCATGATTTTC
ACAACCTGGCAGACGGGTGTGGCGTTTCTCCTGGGAGCCGTTATGAGTGCATCTGCTGGTATCGTGGGAATGA
AGATGGCCACCAGGGCAAACGTCAGGGTGGCGGAGGCTGCAAGAACCACAAAGAAGATAGGTCCCGCTTTGA
AAGTGGCGTATCAAGGCGGAAGCGTTATGGGTCTTTCCGTTGGTGGATTCGCTCTGCTTGGACTCGTCCTGGTG
TATCTGATTTTCGGAAAGTGGATGGGACAGGTGGACAATCTGAACATATACACGAACTGGCTTGGAATAAACTTT
GTTCCTTTTGCGATGACCGTTTCTGGATACGCCCTTGGATGTTCAATCATTGCCATGTTCGATAGGGTCGGTGGA
GGAGTTTACACGAAGGCCGCTGATATGGCAGCCGACCTCGTTGGGAAAACCGAATTGAACCTCCCGGAAGACG
ATCCCAGGAATCCGGCAACGATCGCAGACAATGTGGGAGACAACGTGGGAGACGTTGCGGGGCTCGGAGCGG
ACCTTTTGGAGAGCTTTGTTGGAGCGATAGTTTCTTCCATAATCCTCGCTTCTTACATGTTCCCAATCTACGTTCA
GAAGATAGGTGAGAATCTGGTACATCAGGTACCGAAGGAGACGATACAGGCGCTCATCAGCTATCCTATCTTCT
TTGCTCTTGTTGGTCTTGGCTGTTCGATGCTTGGAATACTCTATGTGATCGTAAAGAAGCCATCTGATAATCCTC
AAAGAGAACTCAACATCAGCCTCTGGACATCCGCGCTGCTCACAGTTGTCCTCACAGCCTTTCTGACGTATTTCT
ATCTGAAAGATCTTCAGGGACTCGATCTAGTCGGATTCCGATTTGGAGCCATCTCACCCTGGTTCTCAGCGATC
ATAGGCATCTTCTCAGGGATTCTCATAGGATTCTGGGCCGAGTACTACACGAGCTATCGCTACAAACCGACCCA
GTTTCTGGGTAAATCTTCTATCGAAGGAACCGGAATGGTCATTTCGAACGGACTTTCGCTTGGTATGAAGAGCGT
GTTTCCTCCCACTCTAACACTTGTTCTTGGTATTCTTTTTGCTGATTATTTTGCGGGTCTCTACGGTGTGGCGATT
GCCGCGCTTGGAATGCTTTCGTTCGTTGCAACATCCGTTTCTGTTGACAGCTACGGTCCCATTGCGGACAACGC
GGGCGGAATAAGCGAAATGTGTGAACTGGATCCCGAAGTCAGGAAGATCACGGATCACCTCGACGCGGTCGGT
AACACCACCGCAGCCATAGGAAAAGGATTCGCTATAGGATCCGCTATCTTTGCAGCGCTTTCGCTCTTTGCGTC
TTACATGTTCTCCCAGATCAGTCCTTCTGATATTGGGAAGCCTCCTTCACTCGTGCTTCTTCTGAACATGCTCGA
TGCGAGGGTCATAGCAGGCGCCCTCCTTGGTGCAGCCATCACATATTACTTCAGTGGATATCTTATTTCGGCAG
TTACCAAGGCAGCTATGAAGATGGTGGATGAGATCAGAAGACAGGCTCGAGAGATCCCAGGACTGCTGGAAGG
TAAGGCCAAGCCTGACTACAACAGGTGTATTGAGATCACAAGCGACAACGCTCTGAAGCAGATGGGTTATCCTG
CGTTCATAGCGATTCTCACACCGCTTGTGACTGGTTTTCTCCTCGGAGCGGAGTTTGTTGGAGGAGTGTTGATA
GGAACGGTCCTCAGTGGTGCGATGCTCGCAATTCTGACGGCCAATTCTGGAGGAGCCTGGGACAACGCGAAGA
AGTACCTGGAAGCGGGTAATTTGGAAGGGTACGGAAAAGGTTCCGAGCCGCACAAAGCGCTGGTTATAGGAGA
CACGGTTGGAGATCCTCTCAAGGACACCGTTGGACCTTCTCTTGACATTCTCATAAAGATCATGTCGGTGGTCTC
GGTGATAGCTGTCTCCATATTCAAACACGTTCACCTGTTCTGATCTAGTTCTAGTTCTAGTCGAGGGAATTGTCA
TTCTAATATTTTATCCACACACACACCTTAAAATTTTTAGATTAAATGGCATCAACTCTTAGCTTCACACACACACA
CACACCGAAGCTGGTTGTTTTATTTGATTTGATATAATTGGTTTCTCTGGATGGTACTTTTTCTTTCTTGGTTATTT
CCTATTTTAAAATATGAAACGCACACAAGTCATAATTATTCTAATAGAGCACAATTCACAACACGCACATTTCAAC
TTTAATATTTTTTTAGAAACACTTTATTTAGTCTAATTCTTAATTTTTAATATATATAATGCACACACACTAATTTATT
CATTAATTTTTTATTGAGTAGGATTTGAAAATATTTGGTATCTTTGCAAGATGTTTGTATAGAGGGACAAAGAATC
GTCTTTATTATGGTCAAGGCTTTACGTCATAATAGTTCCTGCCCAGCTCTTCTATAATACTTTAAAGATCTCTTCTC
GTTTGCTCCATTTGGAAGTCTCGCTTACGTTTATGCGCCCATACAGACACTCAAGATACACACTTACATGAACGT
ATACAAATTTACTAACACTACTTGAAAATATGAACCACAGTACATCATATTAAGACGTAGTATTCGATGATTGAAG
GCCGCCTCCGCGAAATACCTTTACTGATTTTGCCGGTTAATCGCATCGAAATTTCTTCATCACAAGAAAGCAAAC
AAATCGCCAGGCCATTCTACAAGTTTCCTTTTCTTATGAAGATGTAAAAGCTACTAAGGCGTCATTACTCTACGAA
TTCGTAATCATGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGA
AGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGC
TTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGT
ATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAG
CTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGG
CCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAG
CATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCC
TGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGG
GAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGC
TGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGG
TAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGC
TACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAA
GCCAGTTACCTTCGGAAAAARAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTT
TTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTG
ACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCC
TTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTA
ATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATA
ACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTC
CAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTC
CATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGC
CATTGCTACAGGCATCGTSGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAA
GGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGT

AAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGA TGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGC CCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTC GGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGAT CTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGA ATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATT GTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAA AAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCT TTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTG TCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCT GGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATACTAGAAAGTATAGGAACTTCAGAGCG CTTTTGAAAACCAAAAGCGCTCTGAAGACGCACTTTCAAAAAACCAAAAACGCACCGGACTGTAACGAGCTACTA AAATATTGCGAATACCGCTTCCACAAACATTGCTCAAAAGTATCTCTTTGCTATATATCTCTGTGCTATATCCCTAT ATAACCTACCCATCCACCTTTCGCTCCTTGAACTTGCATCTAAACTCGACCTACATTTTTTGTTTATCTAGTATTAC TCTTTAGACAAAAAATTGTAGTAATCATAGGTACAAAACGAAAGAATTTCTACGTATTTGGAATAGAAACTCT
† The original pMP649 backbone sequence into which the *Tm*-PPase gene was inserted into was lost. The sequence listed here originates from sanger sequencing done by former group member Dr. Juho Kellosalo and covers about 8kbp including the protein coding region.

Table S4: Biochemical and biophysical properties of M-PPase constructs

Open reading frame	Molecular weight (Da)	Extinction coefficient (M^{-1} cm^{1})*	Number of tryptophan	Theoretical isoelectric point
pDDGFP2-*Cl*-PPase	101,358.10	72,200	5	6.30
pZH-*Cl*-PPase	120,007.46	118,150	12	6.14
pZH-*Cl*-PPase†	72,912.16	51,800	4	7.13
pMP649-*Pa*-PPase	76,331.11	80,220	7	7.11
pMP649-*Tm*-PPase	78,153.55	80,220	7	6.65

*Assuming all cysteine residues are reduced.
† Data given for the HRV3C cleavage produc.

Supplementary information Chapter 3

Table S5: List of critical residues for M-PPase integrity used as an exclusion set for IMPROvER. This table is adapted from Harborne et al. (173)

Cl-PPase residue	M-PPase study †	Mutation studied	B&W ‡	Role\Effect	Ref *
S173	*Tm*-PPase, AVP1, AVP1, *Sc*-PPase, *Vr*-PPase	M174, E229Q, E229D, K190R, E225D	5.33	facilitates ion release	(121,123,127,128)
L176	*Tm*-PPase, M-PPase, *Sc*-PPase, *Sc*-PPase, *Vr*-PPase	S177, E193A, E193D, T228	5.36	facilitates ion release	(127,121,128,292)
M180	*Tm*-PPase	L181	5.4	prevents back-flow of ions	(127)
R190	*Tm*-PPase, *Rb*-PPase, *Rb*-PPase, *Sc*-PPase, *Sc*-PPase, AVP1, AVP1, *Vr*-PPase, *Vr*-PPase	R191, R176A, R176K, R207A, R207K, R246A, R246K, R242, R242	5.42	coupling of pumping and hydrolysis	(120–122,127,128)
K198	*Tm*-PPase, *Vr*-PPase, *Vr*-PPase	K199, K250A, K250	5.58	coupling of pumping and hydrolysis	(127,128,132)
D231	*Tm*-PPase, *Vr*-PPase, *Vr*-PPase, *Rb*-PPase, *Rb*-PPase, *Sc*-PPase, *Vr*-PPase, *Rb*-PPase, *Rb*-PPase	D232, D283A, D283E, D217A, D217E, D248G, D283, D217A, D217H	-	loose coupling	(119–121,127,128,295)
D235	*Tm*-PPase, *Vr*-PPase, *Vr*-PPase, *Sc*-PPase, *Vr*-PPase	D236, D287A, D287E, D252G, D287	6.43	nucleophile coordination	(119,121,127,128)
D242	*Tm*-PPase, *Sc*-PPase, *Vr*-PPase, *Vr*-PPase, *Vr*-PPase, *Vr*-PPase, *Vr*-PPase	D243, D259G, D294A, D294N, D294E, D294T, D294A, D294	6.5	ion-selectivity	(121,127,128,132)
E245	*Tm*-PPase, *Rb*-PPase, *Rb*-PPase, *Sc*-PPase, *Sc*-PPase, *Sc*-PPase, *Vr*-PPase, ChlPPase, *Bv*-PPase, *Bv*-PPase, *Bv*-PPase	E246, E231Q, E231D, E262A, E262D, E262Q, G297A, E242D, E246A, E246Q, E246D	-	enhanced activity	(120–122,127,292), HS

S253	*Tm*-PPase, *Vr*-PPase, *Vr*-PPase, *Vr*-PPase	S254, A305S, A305, A305	6.61	facilitates ion release	(122,127,128)
A260	*Tm*-PPase, *Vr*-PPase, *Vr*-PPase	Y261, I312A, I312	6.68	facilitates ion release	(122,127,128)
D484	*Tm*-PPase	D458	-	coupling of pumping and hydrolysis	(127)
D445	*Tm*-PPase, *Vr*-PPase, *Vr*-PPase	D465, D507A, D507	11.57	substrate/ product binding	(127,128,132)
D472	*Tm*-PPase, *Vr*-PPase, *Vr*-PPase	N492, N534A, D534	12.43	substrate/ product binding	(127,128,132)
K479	*Tm*-PPase, *Rb*-PPase, *Rb*-PPase, *Rb*-PPase, *Rb*-PPase, *Bv*-PPase, *Bv*-PPase	K499, K469A, K469D, K469R, K469R, K489R, K489A	-	loose coupling/ low affinity	(120,127,295), HS
A482	*Tm*-PPase, *Sc*-PPase, *Sc*-PPase, *Sc*-PPase, *Sc*-PPase, *Sc*-PPase, *Sc*-PPase, *Sc*-PPase, *Sc*-PPase, *Sc*-PPase, *Sc*-PPase, *Sc*-PPase, *Sc*-PPase, *Vr*-PPase	A502, A514S, A514L, A514I, A514M, A514P, A514Y, A514F, A514W, A514D, A514E, A514N, A514H, A514K, A514R, A544	12.53	coupling of pumping and hydrolysis	(125,127,128)
I483	*Tm*-PPase, *Vr*-PPase, *Vr*-PPase	I503, I545A, I545	12.54	coupling of pumping and hydrolysis	(127,128,132)
A486	*Tm*-PPase, *Vr*-PPase	A506, A548	12.57	coupling of pumping and hydrolysis	(127,128)
A487	*Tm*-PPase, *Vr*-PPase	I507, A549	12.58	coupling of pumping and hydrolysis	(127,128)
L493	*Tm*-PPase, *Vr*-PPase, *Vr*-PPase, *Vr*-PPase	L513, L555A, L555D, L555	12.64	prevent back-flow of ions?	(127,128,132)
S496	*Tm*-PPase	S516	12.67	facilitates ion release	(127)
S246	*Tm*-PPase, *Sc*-PPase, *Sc*-PPase, *Sc*-PPase, *Vr*-PPase, *Vr*-PPase, ChlPPase	S247, S263C, S263A, S263E, S298A, S298, S243A	6.54	sodium binding impaired	(121,122,124,127,128,292)
K638	*Tm*-PPase, *Vr*-PPase	K663, K694	15.64	substrate/ product binding	(127,128)
K639	*Tm*-PPase, *Vr*-PPase, *Vr*-PPase	K664, K695A, K695	15.65	substrate/ product binding	(127,128,132)
D661	*Tm*-PPase, *Vr*-PPase, *Vr*-PPase, *Vr*-PPase	D688, D723A, D723E, D723	16.31	substrate/product binding	(119,127,128)
D665	*Tm*-PPase, *Vr*-PPase, *Vr*-PPase, *Vr*-PPase	D692, D727A, D727E, D727	16.35	substrate/ product binding	(119,127,128)
K668	*Tm*-PPase, *Vr*-PPase	K695, K730	16.38	substrate/ product binding	(127,128)
D669	*Tm*-PPase, *Vr*-PPase, *Vr*-PPase, *Vr*-PPase, *Bv*-PPase, *Bv*-PPase	D696, D731A, D731E, D731, D704E, D704A	-	no activity	(119,127,128), HS
N676	*Tm*-PPase, *Vr*-PPase, *Vr*-PPase, ChlPPase	D703, N738A, N738, N677D	16.46	sodium binding impaired	(127,128,132,292)
K680	*Tm*-PPase, *Vr*-PPase, *Vr*-PPase, *Vr*-PPase, *Vr*-PPase, ChlPPase	K707, K742A, K742R, K742M, K742, K681N	-	no activity	(127,128,292)
M684	*Tm*-PPase, *Vr*-PPase, *Vr*-PPase, *Vr*-PPase	V711, V746A, V746D, V746	16.54	prevent back-flow of ions	(127,128,132)
S691	*Tm*-PPase, *Vr*-PPase	S718, P753	16.61	facilitates ion release	(127,128)
F76	AVP1	E119Q	-	loose coupling	(123)
G249	AVP1, AVP1, *Vr*-PPase, *Vr*-PPase, *Vr*-PPase	E305Q, E305D, E301A, E301, E301	6.57	coupling of pumping and hydrolysis	(122,123,128)
E353	AVP1, AVP1, *Rb*-PPase, *Rb*-PPase, *Rb*-PPase	E427Q, E427D, E351D, E351A, E351Q	-	loose coupling/ low affinity	(120,123)
D438	AVP1, AVP1, *Rb*-PPase	D504N, D504E, D428N	-	no activity	(120,123)
T513	AVP1	D573N	-	enhanced activity	(123)
N607	AVP1, *Vr*-PPase	E667Q, E663A	-	loose coupling	(123,132)
F685	AVP1	E751Q	-	loose coupling	(123)
D201	*Vr*-PPase, *Vr*-PPase, *Sc*-PPase, *Vr*-PPase	D253A, D253E, D218G, D253	-	substrate/ product binding	(119,125,128)
V207	*Vr*-PPase	V259A	-	loose coupling/ low affinity	(119)
K209	*Vr*-PPase, *Vr*-PPase, *Vr*-PPase	K261A, K261R, K261	-	stabilising salt bridges	(119,128)

E211	*Vr*-PPase, *Vr*-PPase, *Vr*-PPase, *Vr*-PPase, *Vr*-PPase	E263G, E263A, E263D, E268, E263	-	stabilising salt bridges	(119,128)
D227	*Vr*-PPase, *Vr*-PPase, *Sc*-PPase, *Vr*-PPase	D279A, D279E, D244G, D279	-	substrate/ product binding	(119,121,128)
I252	*Vr*-PPase	C304R	-	loose coupling/ low affinity	(119)
A475	M-PPase		12.46	K^+ dependency	(141)
G478	M-PPase		12.49	K^+ dependency	(141)
V236	*Sc*-PPase	C253A	-	no activity	(124)
T369	*Sc*-PPase	S402C	-	no activity	(124)
E566	*Sc*-PPase	S609C	-	no activity	(124)
C578	*Sc*-PPase	C621A	-	no activity	(124)
S651	*Sc*-PPase	S694C	-	no activity	(124)
T87	M-PPase		3.45	dual pumping signature	(292)
F91	M-PPase		3.49	dual pumping signature	(292)
D143	M-PPase		4.6	dual pumping signature	(292)
R105	*Rb*-PPase	R101K	-	enhanced activity	(120)
G192	*Rb*-PPase	G178A	-	loose coupling/ low affinity	(120)
L387	*Rb*-PPase, AVP1, AVP1	E385Q, K461R, K461A	-	pump-less	(120,122)
E585	*Rb*-PPase, *Rb*-PPase	E584A, E584D	-	loose coupling/ low affinity	(120)
S630	*Rb*-PPase	G637A	-	loose coupling/ low affinity	(120)
T172	*Sc*-PPase	P189L	-	pump-less	(121)
A174	*Sc*-PPase	V191A	-	loose coupling	(121)
T177	*Sc*-PPase	G194Q	-	loose coupling	(121)
F178	*Sc*-PPase	F195L	-	pump-less	(121)
G181	*Sc*-PPase	G198A	-	no activity	(121)
S183	*Sc*-PPase	A200T	-	loose coupling	(121)
M185	*Sc*-PPase	I202T	-	loose coupling	(121)
F188	*Sc*-PPase	F205S	-	loose coupling	(121)
G193	*Sc*-PPase	G210A	-	pump-less	(121)
G194	*Sc*-PPase	G211A	-	pump-less	(121)
D205	*Sc*-PPase, *Vr*-PPase	D222G, D257	-	substrate /product binding	(121,128)
L206	*Sc*-PPase	L223P	-	no activity	(121)
A212	*Sc*-PPase	Q229R	-	loose coupling	(121)
E216	*Sc*-PPase	E233G	-	loose coupling	(121)
I225	*Sc*-PPase	I242T	-	loose coupling	(121)
V229	*Sc*-PPase, *Sc*-PPase	V246I, V246A	-	loose coupling	(121)
M239	*Sc*-PPase	M256T	-	loose coupling	(121)
Y244	*Sc*-PPase, *Vr*-PPase	F261L, F296A	-	loose coupling	(121,122)
G262	*Sc*-PPase, *Vr*-PPase	A278V, S314A	-	low activity	(121,122)
F265	*Sc*-PPase, *Vr*-PPase, *Vr*-PPase	D281G, L317A, L317A	-	low activity	(121,122,132)
P271	*Sc*-PPase	P287A	-	loose coupling	(121)
I277	*Sc*-PPase	I293V	-	loose coupling	(121)
G278	*Sc*-PPase	G294R	-	no activity	(121)
T291	*Sc*-PPase	P307A	-	loose coupling	(121)
N299	*Sc*-PPase	S313E	-	loose coupling	(121)
L306	*Sc*-PPase	R320C	-	loose coupling	(121)
S311	*Sc*-PPase	S325R	-	no activity	(121)
L314	*Sc*-PPase	I328T	-	loose coupling	(121)
T308	*Sc*-PPase	F322L	-	loose coupling	(121)
A318	*Sc*-PPase	L332P	-	pump-less	(121)
L327	*Sc*-PPase	V351A	-	pump-less	(121)
G328	*Sc*-PPase	A357G	-	loose coupling	(121)
Y335	*Sc*-PPase, *Sc*-PPase	L368P, L368A	-	no activity	(121)
G341	*Sc*-PPase	G374A	-	no activity	(121)
A344	*Sc*-PPase	L377P	-	pump-less	(121)
I348	*Sc*-PPase, *Sc*-PPase	I381P, I381A	-	loose coupling	(121)
G349	*Sc*-PPase	Q382R	-	loose coupling	(121)
L243	*Vr*-PPase	L295A	-	loose coupling	(122)
Y247	*Vr*-PPase	Y299A	-	no expression	(122)
V248	*Vr*-PPase, *Bv*-PPase	A300S, C249S	-	enhanced activity	(122), DD
S250	*Vr*-PPase	S302A	-	enhanced activity	(122)
I251	*Vr*-PPase	S303A	-	low activity	(122)
T254	*Vr*-PPase	A306S	-	no expression	(122)
A255	*Vr*-PPase	L307A	-	no expression	(122)

A256	*Vr*-PPase	V308A	-	loose coupling	(122)
L257	*Vr*-PPase	V309A	-	low activity	(122)
A258	*Vr*-PPase	A310S	-	loose coupling	(122)
A261	*Vr*-PPase	S313A	-	low activity	(122)
L263	*Vr*-PPase	F315A	-	low activity	(122)
G264	*Vr*-PPase, *Vr*-PPase	G316A, G316A	-	loose coupling	(122,132)
K266	*Vr*-PPase	N318A	-	no expression	(122)
T157	*Vr*-PPase	D218A	-	loose coupling	(132)
T197	*Vr*-PPase	T249A	-	no activity	(132)
D217	*Vr*-PPase, *Vr*-PPase	D269A, D269	-	substrate/product binding	(128,132)
N228	*Vr*-PPase	N280A	-	low activity	(132)
E698	*Vr*-PPase	E698A	-	loose coupling	(132)
I687	*Vr*-PPase, *Vr*-PPase	L749A, L749D	-	no activity	(132)
Y355	*Sc*-PPase, *Sc*-PPase, *Sc*-PPase, *Sc*-PPase, *Sc*-PPase, *Sc*-PPase	F388Y, F388G, F388P, F388D, F388E, F388K, F388R	-	no expression	(125)
D635	*Vr*-PPase	D691	-	substrate/product binding	(128)
R553	*Vr*-PPase, *Tm*-PPase	R609, R578	13.62	links motion of inner to outer ring	(128)
S485	*Vr*-PPase	S547	-	coordination of water	(128)
M189	*Vr*-PPase	L232	5.4	prevent back-flow of ions	(128)
E500	*Vr*-PPase	R562	12.71	facilitates ion release	(128)
P374	*Rb*-PPase	H372A	-	loose coupling/ low affinity	(120)
E645	*Rb*-PPase	H652A	-	loose coupling	(120)
H654	*Rb*-PPase	H661A	-	loose coupling/ low affinity	(120)
S695	*Rb*-PPase	H702V	-	loose coupling/ low affinity	(120)

† M-PPase abbreviations: AVP1; *Arabidopsis thaliana* (isoform 1), *Bv*-PPase; *Bacteroides vulgatus*, ChlPPase; *Chlorobium limicola*, *Pa*-PPase; *Pyrobaculum aerophilum*, *Rb*-PPase; *Rhodospirillum rubrum*, *Sc*-PPase; *Streptomyces coelicolor*, *Tm*-PPase; *Thermotoga maritima*, *Vr*-PPase; *Vigna radiata*

‡ Ballesteros and Weinstein numbering scheme

* References: DD; undergrad thesis Dovile Dormantaite 2016, HS; undergrad thesis Hannah Shephard 2016

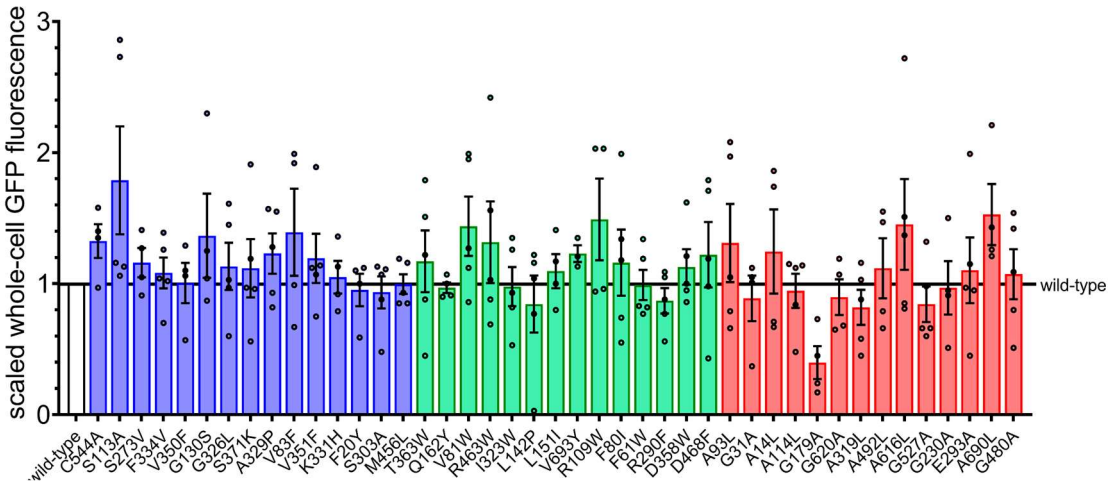


Figure S1: Wild-type and variant *Cl*-PPase GFP-based expression test. Whole-cell GFP fluorescence scaled to wild-type level, which is indicated by a horizontal line. Each individual data point is represented by a circle and error bars denote the SEM. Colouring according to prediction module (deep: blue, model: green, data-driven: red).

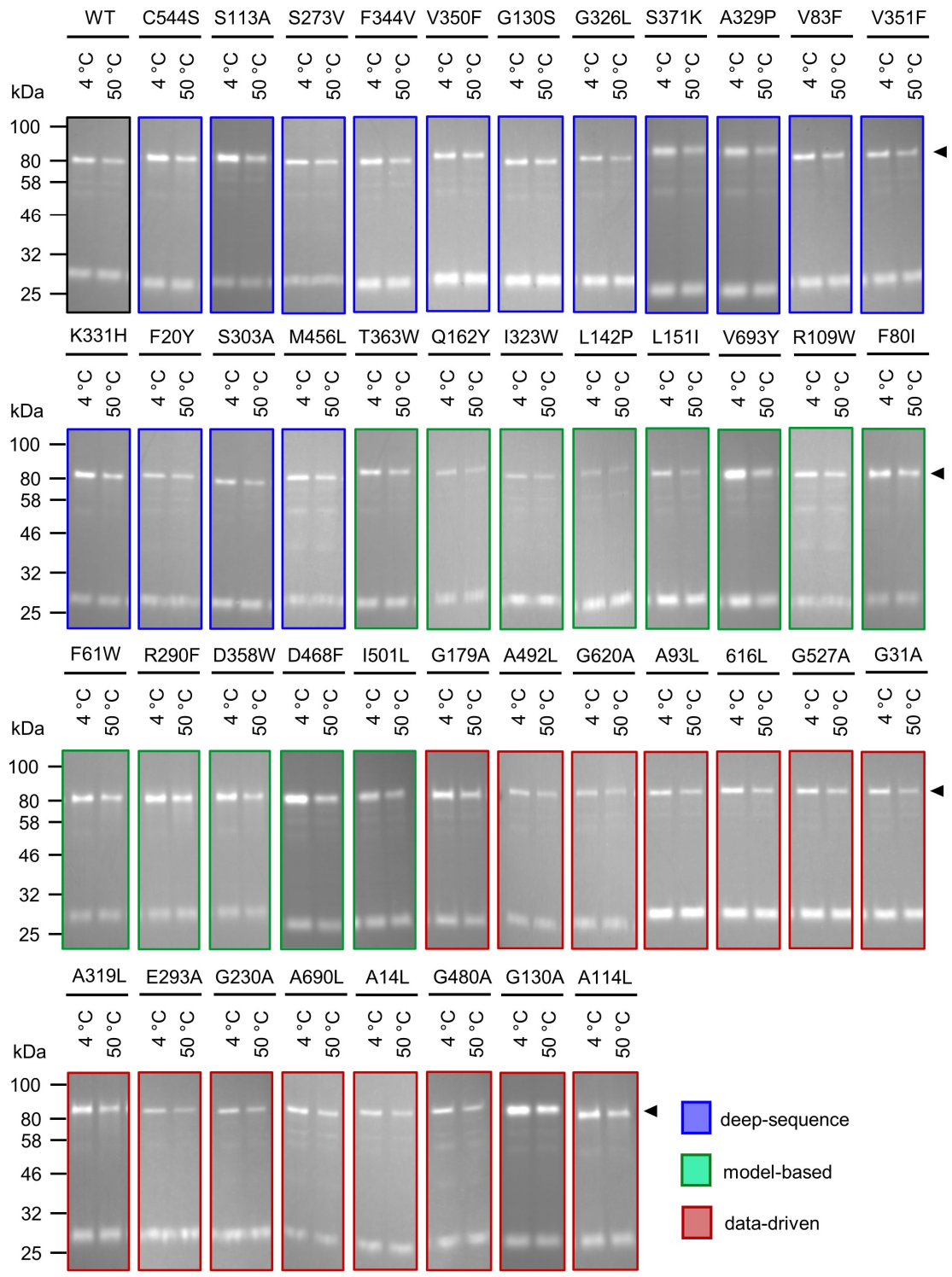


Figure S2: One-temperature thermostability assay of wild-type and variant *Cl*-PPase. Gel images from one representative experimental repeat shown for each variant tested. Colour coding according to prediction module: deep (blue), model (green), data-driven (red). Arrow indicates full-length *Cl*-PPase band. Protein bands were visualised by GFP signal imaging.

Supplementary information Chapter 4

Table S6: Helix by helix comparison of *Pa*-PPase:Mg_5IDP structure to other M-PPase structures.

Rmsd (Å)[ǂ]	*Pa*-PPase:Mg_5IDP *vs.* *							5LZQ *vs.* 4A01*
	5LZQ	4A01	4AV3	4AV6	5GPJ	6AFS	average	
Overall	1.24	1.50	1.67	1.236	1.50	1.49	1.44	0.86
Helix 1	2.16	2.70	2.31	2.23	2.21	2.72	2.39	4.70
Helix 2	1.33	2.29	1.43	1.66	2.41	2.42	1.92	0.86
Helix 3	1.37	2.06	1.93	1.33	1.57	2.06	1.72	4.78
Helix 4	1.18	1.26	1.79	1.58	1.57	1.30	1.45	1.98
Helix 5	1.79	1.61	1.79	1.46	1.09	1.37	1.52	1.41
Helix 6	1.59	1.18	1.32	1.44	0.89	1.10	1.25	1.35
Helix 7	1.84	2.38	2.28	2.07	2.34	2.39	2.22	1.68
Helix 8	1.87	2.09	2.97	2.06	1.81	2.06	2.14	0.96
Helix 9	1.25	1.48	1.85	1.46	1.63	1.48	1.52	1.05
Helix 10	1.20	1.79	1.43	1.09	1.78	1.77	1.51	2.39
Helix 11	1.10	1.43	1.52	1.23	1.34	1.37	1.33	1.92
Helix 12	0.91	0.94	2.77	1.13	2.05	0.94	1.45	0.68
Helix 13[ǂ]	3.01	3.26	1.01	1.03	1.71	3.23	2.21	1.27
Helix 14[ǂ]	2.94	2.73	1.78	1.87	1.85	2.73	2.32	0.70
Helix 15	0.83	0.97	1.30	0.86	1.12	0.95	1.01	0.73
Helix 16	1.44	1.42	1.33	1.26	1.35	1.40	1.37	1.04
Inner ring helices	1.28	1.26	1.67	1.23	1.31	1.19	1.28	-
Outer ring helices	1.82	2.20	1.88	1.64	1.89	2.21	1.82	-

[ǂ] Structural alignment based on Cα atoms of subunit A.
* PDB identifier column header. 5LZQ: *Tm*-PPase:Mg_5IDP, 4AV6: *Tm*-PPase:Mg_4P_{i2}, 5LZR: *Tm*-PPase:Mg_2WO_4, 4AV3: *Tm*-PPase:CaMg, 4A01: *Vr*-PPase:Mg_5IDP, 5GPJ: *Vr*-PPase:Mg_2P_i, 6AFS: *Vr*-PPase: Mg_5P_{i2}
[ǂ] *Pa*-PPase:Mg_5IDP helices cut to length to helices from comparison structures for alignment

Table S7: HELANAL-Plus curvature analysis of helix 5 of *Pa*-PPase:Mg_5IDP.

HELANAL-Plus Parameters[‡]	Helix 5 *Pa*-PPase:Mg_5IDP	Helix 5 *Vr*-PPase:Mg_5IDP
Helix length (residues)	37	37
Average number of residues per turn	3.66	3.69
Average unit hight of helix (Å)	1.5	1.53
Average virtual torsion angle (°)	49.8	49.8
Average bending angle (°)	11.9	11
Maximum bending angle (°)	24.5	23.2
Radius of sphere curvature (Å)	114	76
Rmsd of sphere fit (Å) (rmsdS)	0.247	0.266
Rmsd of linear fit (Å) (rmsdL)	0.154	0.238
Geometry*	linear	kinked

[‡] Definition of all parameters can be found in (191).
* Classification: Linear if, rmsdS > rmsdL and maximum bending angle (MBA) < 20° and if the 20° <MBA< 30°, rmsdS and rmsdL are both <0.14 and 0.16 Å, respectively, and rmsdS > rmsdL. Kinked, if the value of the MBA is >30° or the MBA is between 20° and 30° and rmsd to sphere (rmsdS) and 3D line (rmsdL) fit is more than 0.14 and 0.16 Å, respectively.

Table S8: Comparison of the hydrogen bonding pattern around $S^{5.43}$ and $D^{6.43}$ in IDP-bound structures

Residue from	Helical geometry*		
	Pa-PPase:Mg_5IDP	*Vr*-PPase:Mg_5IDP	*Tm*-PPase:Mg_5IDP
5.37	(α)	3_{10}	3_{10}
5.38	3_{10}	α	3_{10}
5.39	3_{10}	α	(α)
5.40	(α)	α	(α)
5.41	α	α	α
5.42	3_{10}	α	α
5.43	**α**	**α**	**α**
5.43	**3_{10}**	**3_{10}**	**α**
5.44	α	α	α
5.45	(α)	α	α
5.46	α	α	α
6.40	α	(α)	3_{10}
6.41	3_{10}	α	3_{10}
6.42	α	**α**	**α**
6.43	**α**	**(α)**	**(α)**
6.44	**(α)**	**π**	**π**
6.45	**π**	**π**	**π**
6.46	**π**	**α**	**α**
6.47	**α**	3_{10}	π

* Hydrogen bonding from amino nitrogen to upstream carbonyl oxygen of main chain
Entries in parentheses are based on rise per residue as no hydrogen bond (distance > 4 Å) is formed

Table S9: Rotamer options for $K^{12.46}$ in *Pa*-PPase:Mg_5IDP of the backbone-independent Richardson library. Sorting based on their VdW radii overlap to surrounding atoms from lowest to highest.

Rotamer	Chi_1	Chi_2	Chi_3	Chi_4	Sum VdW radii overlap [Å]	Clashes*	Hydrogen bonds
selected	62	-170	68	180	0.00	0	3
1	62	180	68	180	0.62	1	3
2	62	180	-68	180	2.00	2	4
3	63	-178	178	-179	2.55	3	2
4	63	-170	-177	72	2.84	3	1
5	62	180	180	-65	3.03	3	1
6	-70	-179	-66	-64	4.59	4	2
7	-70	-170	-66	-175	5.96	5	3
8	179	59	163	60	7.35	7	0
9	-177	180	68	65	7.46	7	1
10	-59	-69	-176	-70	7.53	8	3
11	-58	-61	-177	-179	8.31	10	2
12	-69	164	62	-179	8.44	7	2
13	180	179	78	179	8.50	9	4
14	-67	-176	174	76	9.04	8	0
15	-177	178	179	180	9.35	9	1
16	-177	68	180	-65	9.62	8	1
17	-62	-68	180	65	9.74	10	1
18	-177	180	171	63	10.01	10	2
19	-90	68	180	180	10.23	10	2
20	179	172	178	-72	10.36	8	1
21	-177	62	173	171	11.62	11	2
22	-67	-179	-179	-63	12.45	11	1
23	-59	-58	-75	-174	12.62	12	1
24	-175	-174	-69	179	12.94	10	2
25	-69	-179	70	67	20.40	12	0
26	-177	180	-68	-65	20.60	13	0
27	-67	176	179	177		7	1

The chi angles translate to m for minus (-60°), t for trans (±180°) and p for plus (+60°) per chi angle in Coot
* VdW radii overlap of ≥ 0.6 Å is classified as a clash

Table S10: Rotamer options for $K^{16.50}$ *Pa*-PPase:Mg_5IDP of the backbone-independent Richardson library. Sorting based on their VdW radii overlap to surrounding atoms from lowest to highest.

Rotamer	Chi_1	Chi_2	Chi_3	Chi_4	Sum VdW radii overlap [Å]	Clashes*	Hydrogen bonds
1 (selected)	-70	-170	-66	-175	0	0	4
2	-69	164	62	-179	0	0	1
3	-69	-179	70	67	0	0	0
4	-177	180	-68	-65	0.83	1	0
5	-175	-174	-69	179	2.17	3	0
6	-70	-179	-66	-64	3.08	3	2
7	-59	-58	-75	-174	3.97	3	1
8	-67	176	179	177	6.41	5	2
9	-177	68	180	-65	8.98	9	1
10	62	180	68	180	9.03	7	4
11	-67	-179	-179	-63	9.1	8	0
12	-67	-176	174	76	10.38	9	2
13	-177	178	179	180	10.63	11	1
14	-177	180	68	65	11.13	7	1
15	179	62	173	171	11.68	12	0
16	-62	-68	180	65	12.41	10	1
17	179	59	163	60	12.62	12	0
18	-177	172	178	-72	12.77	11	1
19	-177	180	171	63	12.85	12	0
20	-58	-61	-177	-179	13.73	10	2
21	-59	-69	-176	-70	15.19	11	0
22	180	179	78	179	15.78	11	0
23	63	-178	178	-179	17.19	14	1
24	62	180	180	-65	19.7	15	1
25	63	-170	-177	72	20.33	15	1
26	-177	180	-68	-65	21.22	13	0
27	-90	68	180	180	21.52	15	1
28	62	180	-68	180	22.61	15	1

The chi angles translate to m for minus (-60°), t for trans (±180°) and p for plus (+60°) per chi angle in Coot

* VdW radii overlap of ≥ 0.6 Å is classified as a clash

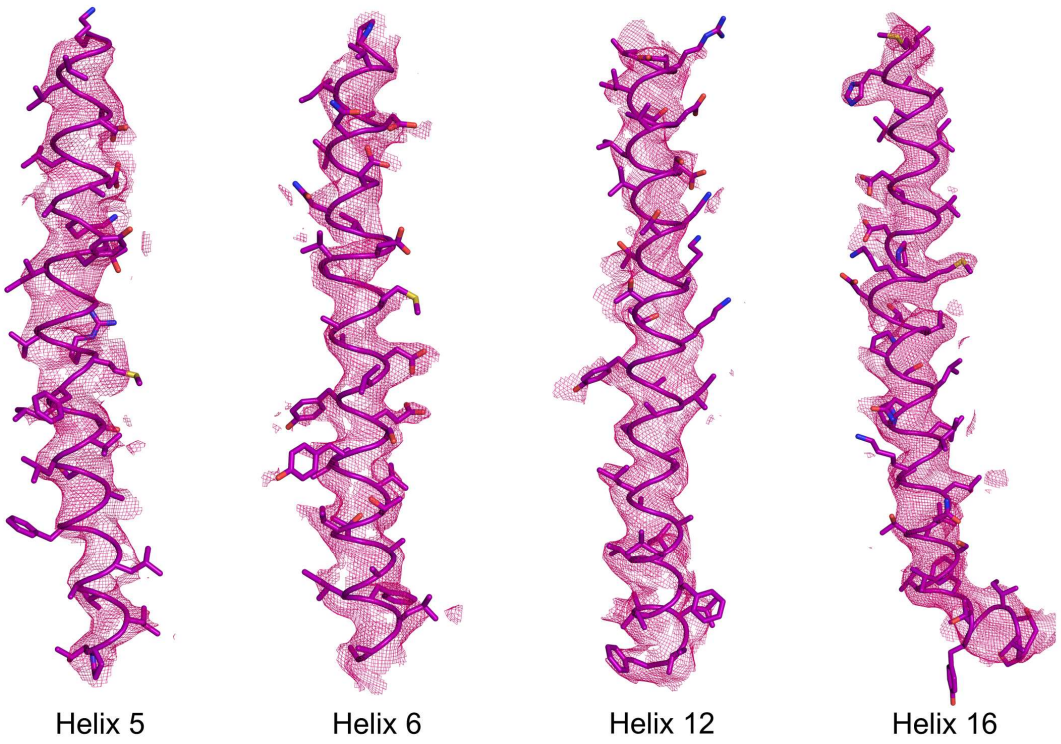


Figure S3: Representative $2F_o$-F_c electron density maps of *Pa*-PPase:Mg_5IDP transmembrane helices.

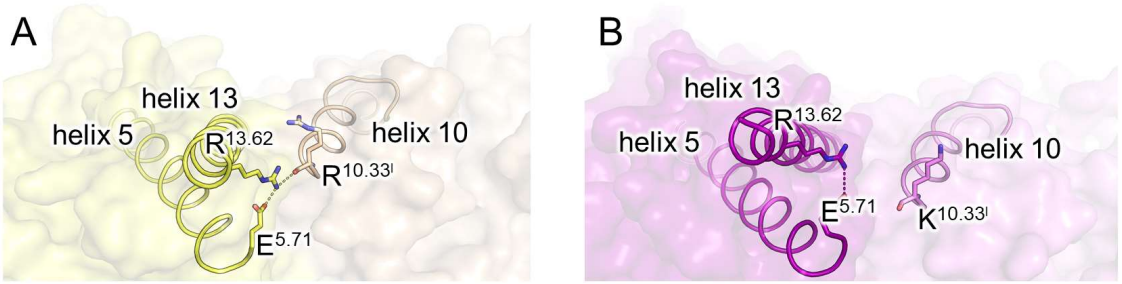


Figure S4: Orientation of helix 5, helix 13 and helix 10 (opposing subunit) in different IDP-bound M-PPase structures. (A) Orientation shown in *Vr*-PPase:Mg_5IDP. (B) Orientation shown in *Pa*-PPase:Mg_5IDP. Salt bridge network interactions of $E^{5.71}$-$R^{13.62}$-R/$K^{10.33}$ are represented by dashed lines in both panels.

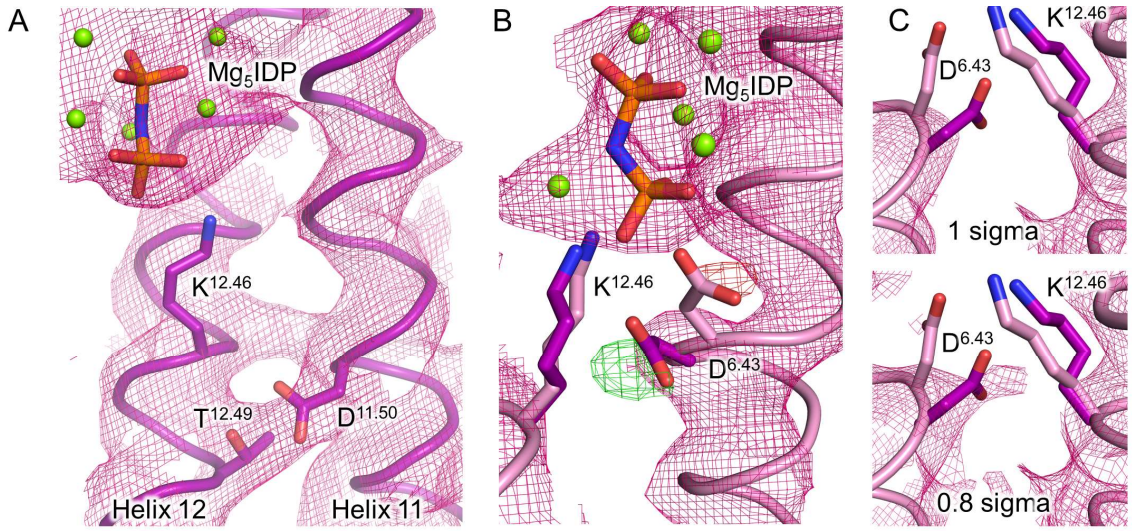


Figure S5: Electron density maps of the $K^+/K^{12.46}$ cationic centre and nearby residues in *Pa*-PPase:Mg_5IDP (purple). (A) $2F_o$-F_c map for key residues involved in K^+-independence. (B) $2F_o$-F_c map and F_o-F_c difference map for model with $D^{6.43}$ orientation built as in high-resolution *Vr*-PPase:Mg_5IDP structure (pink). Additionally, the $D^{6.43}$ and $K^{12.46}$ orientation of the final *Pa*-PPase:Mg_5IDP model are shown (purple). The $D^{6.43}$ orientation of the final model fits the positive (green) and negative (red) F_o-F_c difference map peaks better. (C). Close-up view of $D^{6.43}$ orientation built as seen in high-resolution *Vr*-PPase:Mg_5IDP structure (pink). Additionally, the $D^{6.43}$ and $K^{12.46}$ orientation of the final model are shown in purple. Upper panel shows $2F_o$-F_c density at 1 sigma and lower panel shows $2F_o$-F_c density at 0.8 sigma. The $D^{6.43}$ orientation of the final model fist the 0.8 sigma $2F_o$-F_c map better.

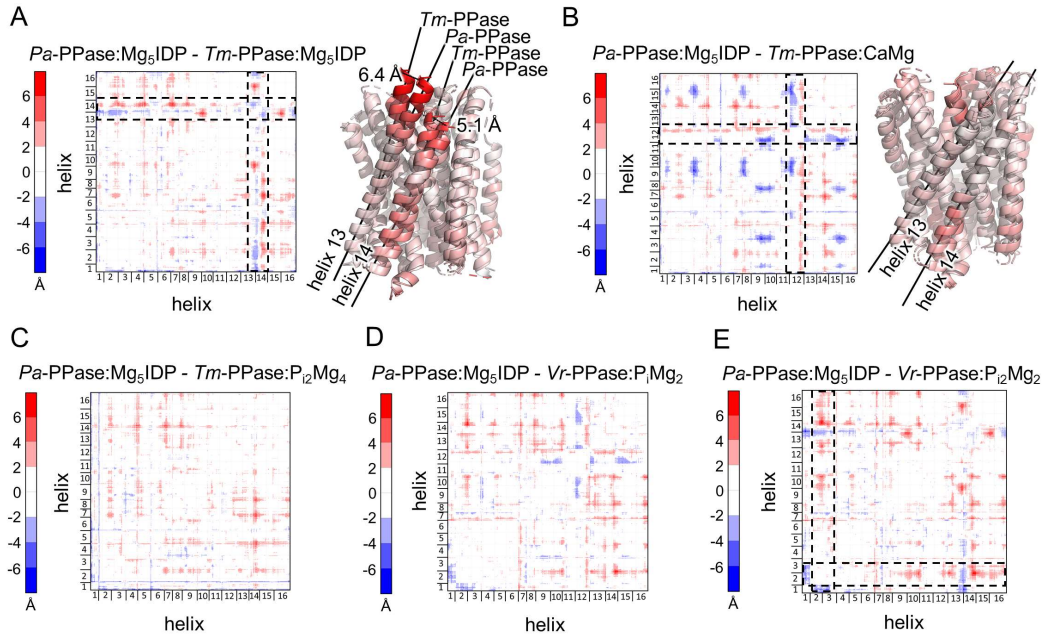


Figure S6: Structural comparison *Pa*-PPase:Mg_5IDP to other M-PPases. (D-H) DiDiMa of *Pa*-PPase:Mg_5IDP minus (D) *Tm*-PPase:Mg_5IDP, (E) *Tm*-PPase:CaMg, (F) *Tm*-PPase: Mg_4P_{i2}, (G) *Vr*-PPase: Mg_2P_i or (H) *Vr*-PPase:$P_{i2}Mg_2$. Major structural changes are highlighted by dashed boxes. Inter atom distances shorter or longer in *Pa*-PPase than in other M-PPases are indicated by blue or red colour, respectively, in the matrix. Right panels next to DiDiMa show the M-PPase monomer with helices coloured by their rmsd according to the DiDiMa scale. The biggest structural changes are highlighted in the structure with respective distance.

Supplementary information Chapter 5

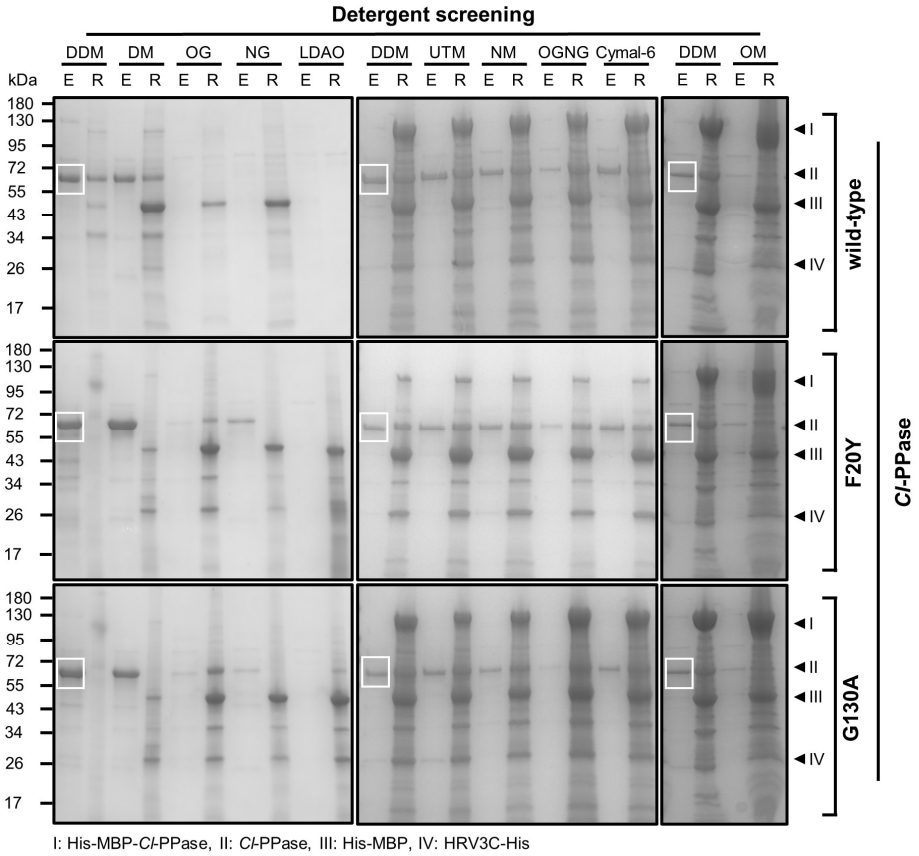


Figure S7: *Cl*-PPase detergent screening. SDS-PAGE analysis (Coomassie stain) of IMAC elution fraction (E) and resuspended resin (R) from small-scale purification of wild-type and variant *Cl*-PPase in different detergents. Purification yields were normalised to the protein amount obtained in DDM after densitometric analysis of *Cl*-PPase bands (II) in ImageJ. DDM reference bands are highlighted by a white box. Other bands of interest are indicated by black arrows with corresponding labels shown below the panel.

Supplementary information Chapter 6

Table S11: X-ray data collection and refinement statistics of time-resolved (0-10 s) *Tm*-PPase structures from combined datasets.

Data collection	**t = 0-300 s**	**t = 0 s**	**t = 1 s**	**t = 2 s**	**t = 5 s**
Dataset	combined	combined	-	combined	combined
Space group	$P2_1$	$P2_1$	-	$P2_1$	$P2_1$
Cell dimensions			-		
a, *b*, *c* (Å)	84.0, 110.2, 108.0	84.0, 110.3, 107.8	-	84.1, 109.9, 107.8	84.0, 110.1, 106.9
α, β, γ (°)	90.0, 107.9, 90.0	90.0, 107.9, 90.0	-	90.0, 108.1, 90.0	90.0, 108.1, 90.0
Source	DESY: P14-I	DESY: P14-I	-	DESY: P14-I	DESY: P14-I
Wavelength (Å)	0.967	0.967	-	0.967	0.967
Resolution (Å)	75.12-2.52 (2.81-2.51)	75.26-2.69 (3.10-2.69)	-	102.45-3.22 (3.31-3.22)	75.11-3.05 (3.22 – 3.05)
Overall (Å)	2.51	2.69	-	3.22	3.05
along h axis	2.51	2.69	-	3.22	3.05
along k axis	2.89	3.32	-	3.22	3.19
along l axis	3.34	3.78	-	3.55	3.69
Measured reflections	2202989 (34742)	351718 (13851)	-	534299 (4693)	177489 (2724)
Unique reflections	39895 (1995)	28084 (1404)	-	27526(1374)	27525 (1375)
Completeness (%)	94.3	92.8	-	96.4	94.0
$CC_{1/2}$	0.995	0.988	-	0.993	0.986
Mean I/σ(I)	13.1	8.1	-	8.6	5.6
Multiplicity	55.2	12.9	-	19.4	6.4
B-factors (Å^2)	56.97	84.85	-	84.97	91.00
R_{merge}	0.278 (4.325)	0.166 (1.048)	-	0.224 (0.335)	0.192 (0.262)
R_{meas}	0.281 (4.459)	0.173 (1.107)	-	0.229 (0.398)	0.209 (0.354)
R_{pim}	0.051 (1.573)	0.048 (0.352)	-	0.049 (0.213)	0.079 (0.263)
Refinement	**t = 0-300 s**	**t = 0 s**	**t = 1 s**	**t = 2 s**	**t = 5 s**
Molecular replacement search model	*Tm*-PPase: CaMg	*Tm*-PPase: 0-300s	-	*Tm*-PPase: 0-300s	*Tm*-PPase: 0-300s
Active site modelled	empty	empty	-	empty	empty
R_{work}(%)/R_{free}(%)	22.45/24.71	23.51/25.06	-	23.80/26.63	
No. of atoms	10597	10514	-	10514	10514
Protein	10453	10475	-	10475	10475
Ligands/Lipids	91	39	-	0	0
Water	53	0	-	0	0
No. of chains (ASU)	2	2	-	2	2
B-factors (Å^2)	75.71	100.19	-	94.73	97.40
Protein	76.06	100.33	-	94.85	97.54
Ligands/Ions	50.66	64.08	-	61.13	58.84
R. M. S. Deviations			-		
Bond lengths (Å)	0.003	0.003	-	0.003	0.002
Bond angle (°)	0.48	0.51	-	0.50	0.41
Ramachandran statistics (%)			-		
Favoured	98.03	97.89	-	98.31	98.17
Allowed	1.97	2.11	-	1.69	1.83
Outliers	0.00	0.00	-	0.00	0.00

Statistics for the highest-resolution shell are shown in parentheses.

Table S12: X-ray data collection and refinement statistics of time-resolved (60-3600 s) *Tm*-PPase structures from combined datasets.

Data collection	**t = 10 s**	**t = 60 s**	**t = 300 s**	**t = 600 s**	**t = 3600 s**
Dataset	combined	combined	combined	combined	combined
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Cell dimensions					
a, *b*, *c* (Å)	83.9, 110.1, 106.7	84.0, 109.9, 106.9	83.9, 110.6, 106.1	83.3, 111.6, 106.2	83.1, 111.1, 105.4
α, β, γ (°)	90.0, 108.0, 90.0	90.0, 108.1, 90.0	90.0, 108.2, 90.0	90.0, 109.1, 90.0	90.0, 108.9, 90.0
Source	DESY: P14-I	DESY: P14-I	DESY: P14-I	DESY: P14-I	DESY: P14-I
Wavelength (Å)	0.967	0.967	0.967	0.967	0.967
Resolution (Å)	79.81-2.78 (2.87-2.78)	101.61 -2.67) (3.07-2-67)	100.81-3.97 (4.09-3.97)	100.4-3.84 (4.17-3.84)	99.74-4.53 (5.22-4.53)
Overall (Å)	2.78	2.67	3.97	3.84	4.53
along h axis	2.78	2.67	3.97	3.86	4.53
along k axis	3.78	3.14	3.97	3.84	5.40
along l axis	3.25	3.65	4.51	4.71	6.00
Measured reflections	306850 (5652)	172847 (5898)	90019 (2490)	85342 (2342)	43682 (2211)
Unique reflections	39499 (1975)	30244 (1512)	14163 (708)	13078 (654)	6063 (303)
Completeness (%)	95.1	92.6	95.7	90.7	87.8
$CC_{1/2}$	0.986	0.980	0.975	0.965	0.995
Mean I/σ(I)	7.6	12.8	6.1	4.4	10.1
Multiplicity	7.8	5.7	6.4	6.5	7.2
B-factors (Å^2)	62.40	93.32	88.03	210.38	286.11
R_{merge}	0.124 (0.341)	0.124 (0.607)	0.187 (0.589)	0.165 (0.779)	0.066 (1.307)
R_{meas}	0.132 (0.420)	0.140 (0.706)	0.203 (0.697)	0.179 (0.918)	0.072 (1.409)
R_{pim}	0.043 (0.240)	0.062 (0.355)	0.079 (0.368)	0.067 (0.479)	0.027 (0.519)
Refinement	**t = 10 s**	**t = 60 s**	**t = 300 s**	**t = 600 s**	**t = 3600 s**
Molecular replacement search model	*Tm*-PPase: CaMg	*Tm*-PPase: 0-300s	*Tm*-PPase: CaMg	*Tm*-PPase: 0-300s	*Tm*-PPase: 0-300s
Active site modelled	empty	empty	empty	empty	empty
R_{work}(%)/R_{free}(%)	22.44/25.00	21.76/23.76	22.79/24.94	29.77/32.82	31.57/36.96
No. of atoms	9965	10514	9965	10514	10475
Protein	9965	10475	9965	10475	10475
Ligands/Lipids	0	0	9	9	0
Water	0	0	0	0	0
No. of chains (ASU)	2	2	2	2	2
B-factors (Å^2)	77.02	107.95	94.34	215.75	366.11
Protein	77.02.	108.08	94.34	215.84	366.11
Ligands/Ions	-	73.69	-	192.68	-
R. M. S. Deviations					
Bond lengths (Å)	0.009	0.002	0.010	0.002	0.004
Bond angle (°)	1.03	0.44	1.19	0.42	0.62
Ramachandran statistics (%)					
Favoured	95.16	98.10	93.94	97.68	97.96
Allowed	4.19.	1.90	4.84	2.32	1.97
Outliers	0.65	0.00	1.23	0.47	0.07

Statistics for the highest-resolution shell are shown in parentheses.

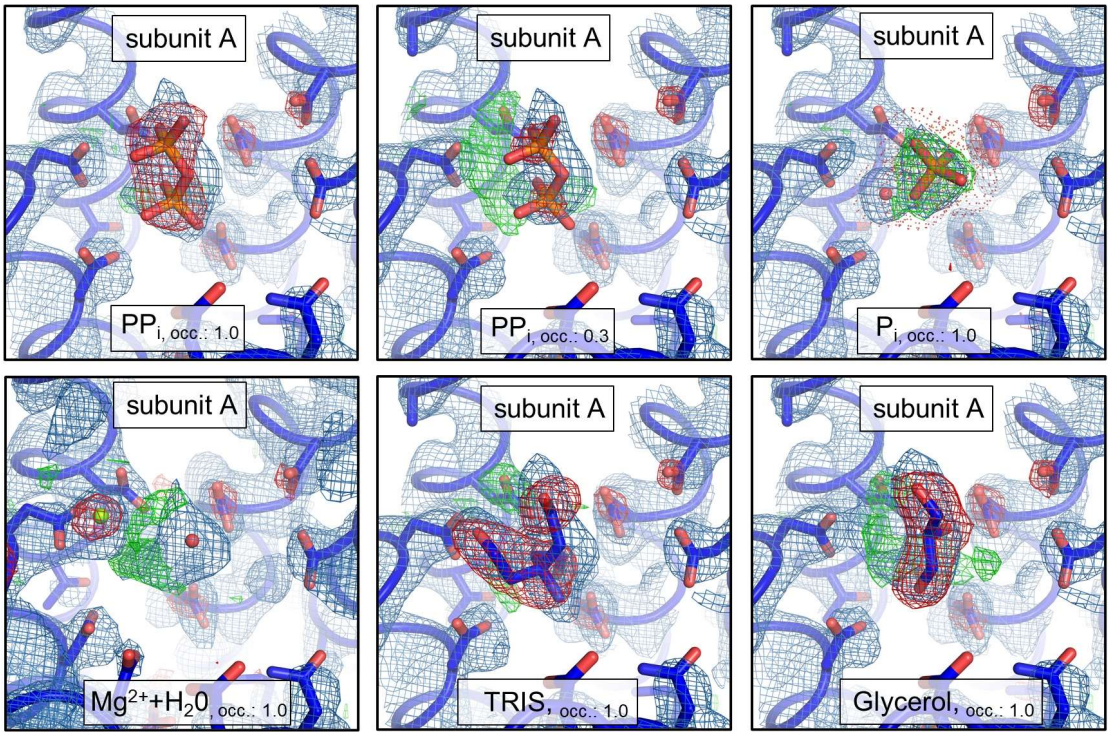


Figure S8: *Tm*-PPase active site of Cluster A. Cluster A includes data from crystals soaked for 0-300 seconds in Na^+-trigger solution and shows an improved $2F_o$-F_c map quality. The panels show the $2F_o$-F_c and F_o-F_c density after refinement with hetero atoms modelled into the active site of subunit A. Water is displayed as red sphere and Mg^{2+} as green sphere. The ligand identity and occupancy (occ) are shown in the lower section of each panel. If not stated otherwise, the $2F_o$-F_c density is contoured at 1 σ and the F_o-F_c density is contoured at 3 σ. PP_i/P_i orientations modelled with guidance of *Tm*/*Vr*-PPase:Mg_5IDP structures.

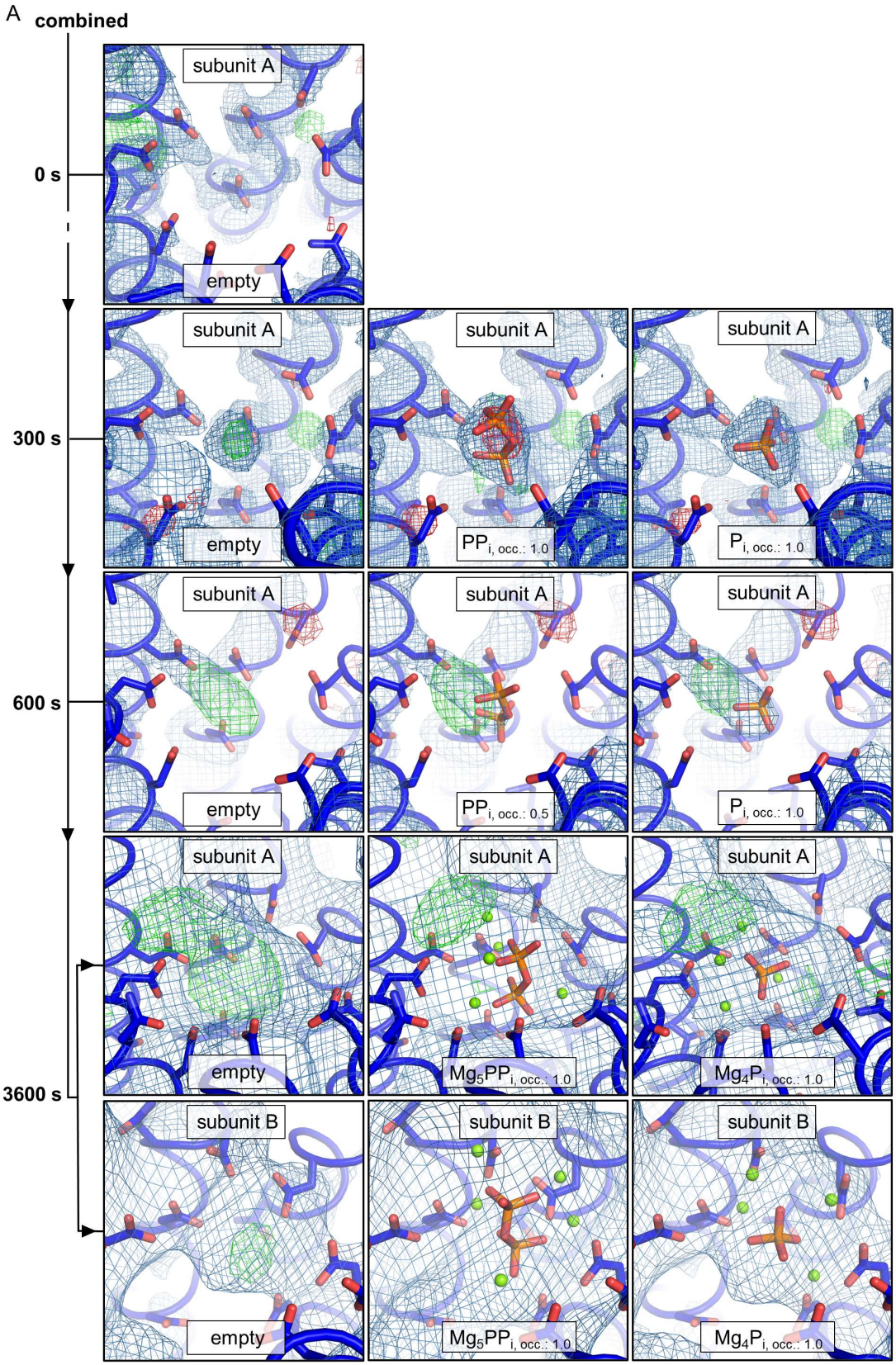


Figure S9: Enzymatic core regions of time-resolved *Tm*-PPase structures obtained from combined datasets. Active site at t=0, 300, 600 and 3600 seconds shown with $2F_o$-F_c density (blue) and F_o-F_c density (red/green). Where hetero atoms were modelled into the active site the resulting F_o-F_c (red/green) after refinement is displayed. The ligand identity and occupancy (occ) are displayed in the lower section of each panel. If not stated otherwise, the $2F_o$-F_c density is contoured at 1 σ and the F_o-F_c density is contoured at 3 σ. PP_i/P_i orientations modelled with guidance of *Tm*/*Vr*-PPase:Mg_5IDP structures.

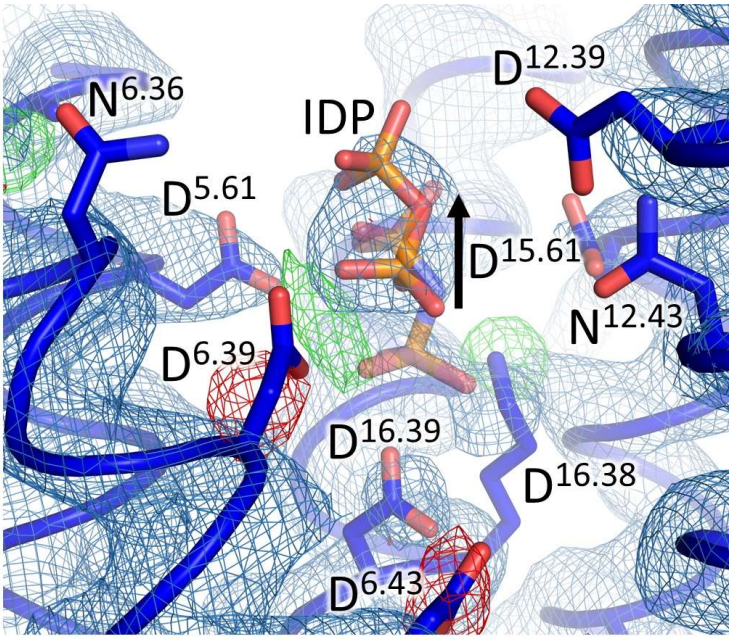


Figure S10: Comparison of IDP binding to active site. Active site of combined 300 seconds structure of *Tm*-PPase with $2F_o$-F_c density and F_o-F_c density shown. IDP orientation in *Vr*-PPase:Mg_5IDP structure is modelled into the active site as well (transparent). Black arrow indicates 3 Å difference between PP_i and IDP binding position.

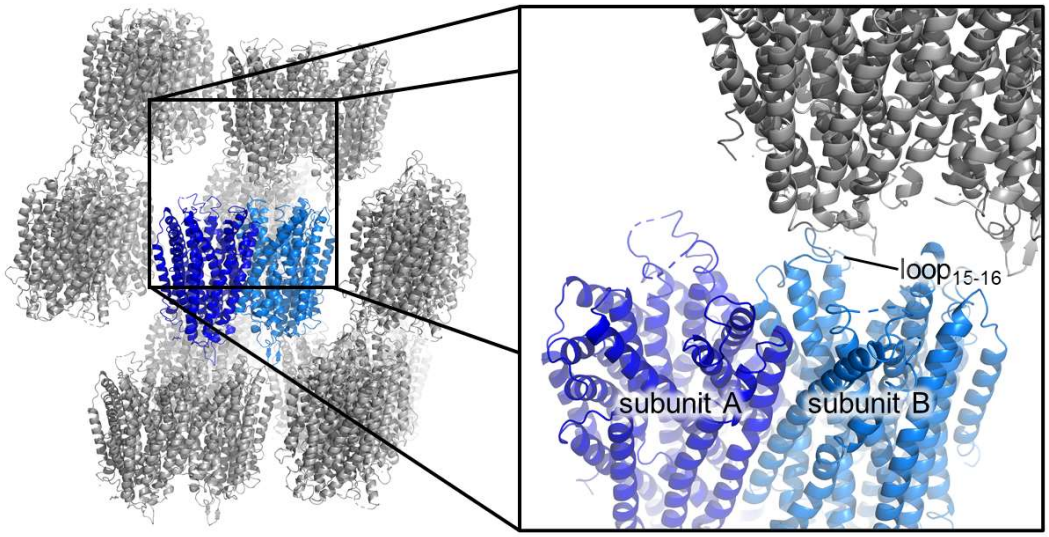


Figure S11: Crystal packing of *Tm*-PPase reference structure prior to reaction initiation by addition of Na^+. The subunits of homodimeric *Tm*-PPase are coloured in different shades of blue and symmetry mates within 20 Å are shown in grey. The black box shows a magnification of crystal contacts near the active site of subunit B, which involves $loop_{15\text{-}16}$. Access to the active site of subunit B does not seem to be affected as indicated by the occurrence of positive F_o-F_c density after 3600 seconds of soaking.