UNIVERSITY OF LEEDS

The Structure and Function of Membrane-Integral Pyrophosphatases and a Mechanosensitive Ion Channel as Studied Through Computational Techniques

Alexandra Oban Morrison Holmes

School of Biomedical Sciences

Faculty of Biological Sciences

The University of Leeds

October 2022

Submitted in accordance with the requirements for the degree of *Doctor of Philosophy*

Intellectual Property and Publication Statement

I confirm that the work submitted is my own, except where work which has formed part of jointly authored publications has been included. My contribution to jointly authored work has been explicitly indicated below. I confirm that appropriate credit has been given within this thesis where reference has been made to the work of others.

- The chapter of my thesis titled “Introduction” includes information from the following co-authored publication:

  Title: **The Function of Membrane Integral Pyrophosphatases From Whole Organism to Single Molecule**
  Journal: **Frontiers in Molecular Biosciences** (doi: 10.3389/fmolb.2019.00132)
  Date Published: 22nd November 2019
  Authors: **Alexandra O. M. Holmes**, Antreas C. Kalli and Adrian Goldman

  My Contribution: Conceptualisation, writing (review and editing), investigation and writing (original draft preparation)

- The chapter of my thesis titled “mPPases Create a Conserved Anionic Membrane Fingerprint as Identified *via* Multi-Scale Simulations” includes information from the following co-authored publication:

  Title: **mPPases Create a Conserved Anionic Membrane Fingerprint as Identified *via* Multi-Scale Simulations**
  Journal: **PLOS Computational Biology** (doi: 10.1371/journal.pcbi.1010578)
  Date Published: 3rd October 2022
  Authors: **Alexandra O. M. Holmes**, Adrian Goldman and Antreas C. Kalli

  My Contribution: Conceptualisation, data curation, formal analysis, investigation, visualisation, writing (original draft) and writing (review and editing)

- The chapter of my thesis titled “Atomistic Investigations of the mPPase Catalytic Cycle” includes data produced in collaboration with other researchers.

  Specifically, the PELDOR measurements were produced by Dr Christos Pliotas, and the *in vitro* performance of compounds was generated by Dr Keni Vidilaseris and presented in the section titled “Performance of the Identified Inhibitors *in vitro*”. All other analyses, data and interpretations were generated by me.

Acknowledgements

In the words of Carmen Maria Machado[1], "it turns out that when you publish a debut book, you have an impossible task: not just thanking the people who directly influenced this particular title, but thanking everyone who has ever had a hand in your becoming a writer." This thesis isn't a debut book – it has far too many plot holes and superfluous detail – but it does mark the end of a particular chapter in my life, and as such I need to thank everyone who has had a hand in me becoming the scientist I am today.

I'd like to start by thanking the member of my supervision team who's stuck through the last four years with me and genuinely kept this entire project on track, Dr Antreas Kalli. Without your knowledge, guidance and kindness, I don't think this work would have been even a fraction of what it is today. You've encouraged me at every turn and supported me to get the most out of the last 4 years, thank you. To the supervisors and teams that I only got to spend a brief amount of time: Prof Colin Fishwick, Dr Martin McPhillie and the entire Fishwick Research Group, thank you for teaching me all I know about *in silico* chemistry. I'd like to thank Prof David Beech and Dr Jian Shi for joining my supervision team mid-way through my PhD. I can imagine how difficult it can be taking on a student part way through their studies, but you have always taken the time to provide guidance and never made me feel like anything less than a valued member of your teams.

I've had the pleasure of being part of so many research groups during the last 4 years. From the entirety of level 6 of the Astbury Building who let me have far too many multi-hour lunch breaks, to the PIEZO group letting me have far too many multi-hour coffee breaks. Thank you to each and every one of you past and present, you've all had a role in this work, but I have to give special commendations to Dr Jamie Henderson (for teaching me everything I know – none of it came in handy, but thanks for stepping up), Dr Brendan Farrell, Dr Claudia Zilian, Dr Jannik Strauss and Dr Jess Boakes (who all tried their absolute best to talk me out of doing this), Chrysa Soteriou (for always being up for a coffee and a catch-up), Dr Chaline Fagnan (who knew there were so many ways a simulation could go wrong?), Dr Marj Debant, Dr Fiona Bartoli and Irish (for all your thesis writing support (one of you being more hands/paws on than the others) and getting me out into nature whenever possible), Dr Chew Cheng, Dr Leander Stewart and Dr Laeticia Lichtenstein (for Friday mornings well spent) and Dr Euli Chuntharpursat and Dr Greg Parsonage (for dragging me kicking and screaming into the labs, then letting me out again).

I also had the support of people throughout and even beyond the university, and I'd like to thank those people for being a balm to my various stresses. Katie Smith for distracting me exactly when I needed it (and especially when I didn't), what will you do without our impromptu book club and my corporeal presence in the office? Maria Nikolova and Philippa Malko, thank you for all those coffee mornings, rants, board game clubs, office Christmas parties, free cinema trips and #HikesWithDykes. I'm glad the two people I met in the first week of my PhD turned out to be the most difficult to get rid of, although maybe I never picked up on it being the other way around... Roz Ffrench, where would I be without you? You've checked-in everyday, you've sent me a copious amount of coffee that's fuelled pretty much all of this, you've provided all of the pet content I could need (but I would never say no to more). Thank you thank you thank you. Shauni Sanderson and Alexa Ruppertsberg, I need to thank you both for being my own personal cheerleaders in the Public Engagement Team. Speaking of public engagement, I'd like to thank the members of the 2019, 2020, 2021 and 2022 Pint of Science Teams I had the joy of working with, getting to discuss science outside of the university left me rejuvenated and motivated to carry on with the work that follows. Thank you Dr Affelia Wibisono for writing a science show with a novice like me, and then for not getting cold feet when we actually had to do it! There are so many other people I could (and should) include here. If you notice you're missing, please know I'm grateful, just very forgetful.

Finally, this acknowledgement section would be entirely remiss without my profound and heartfelt thanks to my family for being the best support I could have asked for. Leslie and Rupert, I'm so sorry for all the tearful phone calls at inopportune moments, threats of dropping out and requests (demands) for early morning cups of coffee. From being the kid who struggled to copy from the class board and learnt my times tables in the bath, to soon becoming a doctor; who'd have thought letting me watch all those hours of Animal Planet and going to the Cheltenham Science Festival would have culminated in me doing a PhD? Thanks for parenting me through this. Olivia, I'm so sorry for stealing your books, clothes and thunder by cementing myself as the gifted sibling (at least on paper).

That leaves me to now poorly paraphrase Ray Bradbury[2]: So here, after 4 years, is my PhD thesis. I didn't know what I was doing, but I'm glad that it was done.

[1]Carmen Maria Machado. 2017. Her Body and Other Parties. Acknowledgements. p.243.

[2]Ray Bradbury. 2003. Fahrenheit 451. Introduction to the 50th Anniversary Ed. p.6.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

APBS	Adaptive Poisson-Boltzmann Solver	MPH	Membrane Parallel Helix
CG	Coarse-grained	mPPase	Membrane-Integral Pyrophosphatase
CHARMM	Chemistry at Harvard Macromolecular Mechanics	PDB	Protein Databank
CHOL	Cholesterol	PELDOR	Pulsed Electron-Electron Double Resonance
C_α	Carbon – alpha	POP2/PIP_2	Phosphatidylinositol Bisphosphate
DPCE	Ceramide Hexoside	POPA	1-Palmitoyl-2-Oleoyl Phosphatidic Acid
DPSM	Sphingomyelin	POPC	1-Palmitoyl-2-Oleoyl Phosphatidylcholine
EM	Electron Microscopy	POPE	1-Palmitoyl-2-Oleoyl Phosphatidylethanolamine
ETI	Etidronic Acid	POPG	1-Palmitoyl-2-Oleoyl Phosphatidylglycerol
FRET	Förster Resonance Energy Transfer	POPS	1-Palmitoyl-2-Oleoyl Phosphatidylserine
GROMACS	Groningen MAchine for Chemical Simulation	PPase	Pyrophosphatase
GUI	Graphical user Interface	PP_i	Inorganic Pyrophosphate
HBU	Helical Bundle Unit	RMSD	Root Mean Squared Deviation
IDP	Imidodiphosphate	RMSF	Root Mean Squared Fluctuation
Insane	INsert membrane	TMH	Transmembrane Helix
LINCS	Linear Constrains Solver	WT	Wild type
MD	Molecular Dynamics		

1. Abstract

To study how membrane proteins perform their functions at a molecular level, and how this is modulated by interactions with lipids and the environment, I selected two protein families for simulation to provide new understanding and ideas.

Membrane-integral pyrophosphatases (mPPases) couple pyrophosphate hydrolysis to cation translocation. Multi-scale simulations demonstrated anionic lipids binding at the mPPase interface and the distal ends of the protein. These sites were composed of positive residues on helices that are implicated in function and stability, suggesting that lipids are important for mPPase activity. The change in dynamics between catalytic states was investigated and, despite no evidence of asymmetry, several mechanistic insights were gained. Such as the mode of sodium ion binding through coordination by the $D^{6.50}/D^{16.46}$ pair to pass $R^{5.50}$, and the mechanism of 5-6 loop closure by forming a hydrophobic interface with the 13-14 loop. Post-simulation structures were used for *in silico* chemical screening and the identification of compounds with improved performance.

Sequential removal of the helical bundle units (HBUs) from mechanosensitive ion channel PIEZO1 uncovered the interplay between blade structure and the membrane. This revealed that the distal HBUs point downwards, in contrast to previous structural predictions, and maintain dome depth at ~5.4 nm. Several points of flexibility along the blade length, modulated by membrane parallel helices (MPH), allowed the formation of three different blade conformations. These data suggested the importance of the MPH and blade structure in controlling dome depth, dynamics and specific interactions with lipids.

By investigating similar findings, such as specific protein-lipid interactions and protein dynamics, in dissimilar proteins, I have generated understanding of their structure, function and their modulation by the membrane, beyond what has been gained through other techniques. This may provide insights into comparable features in other proteins, which could be exploited for stability, functional and structural studies.

2. Introduction

This chapter is partially based on the following publication cited in this thesis as (Holmes *et al*., 2019) Holmes *et al*., The Function of Membrane Integral Pyrophosphatases From Whole Organism to Single Molecule. Front Mol Bio. 2019. doi: 10.3389/fmolb.2019.00132.

2.1 General Introduction to Membrane Proteins

For cells to carry out their varied and various functions they require tools, some of which come in the form of proteins, with which to sense, influence and generally interact with the world around them. The major membrane protein functions, are to sense and report signals from one side of the membrane to the other, facilitate the passage of molecules and ions across the semipermeable barrier, or perform enzymatic tasks at the membrane. Around 30% of the human proteome represents membrane proteins (Almén *et al*., 2009; Bill *et al*., 2011), and they are some of the most physiologically relevant molecules for the function of all organisms, as they typically straddle the barrier between cellular compartments and the cell with its surroundings (Lee *et al*., 2016a). Due to their importance, dysregulation of their production, trafficking, folding, which typically occurs at the membrane (Skach, 2009), or activity can be disastrous for normal cellular physiology. As such, they represent a large proportion of drug targets, in 2006 this figure was estimated at 60% (Overington *et al*., 2006), but likely in the last 16 years this number has increased as more membrane proteins become characterised and amenable to study.

Despite their clinical and functional importance, membrane proteins have been historically neglected in the scientific literature and structure determination, as in 2016 only ~2% of the high-resolution structures in the Protein DataBank are membrane proteins (Lee *et al*., 2016a), and in the subsequent 4 years this number has only increased to ~3% (Kwan *et al*., 2020). As of 2023, this figure now stands at ~3.4% (https://blanco.biomol.uci.edu/mpstruc/), demonstrating the difficulties in gaining membrane protein structures. This is due to a number of barriers to their study that make them difficult to work with in laboratory-settings (Bill *et al*., 2011). They are typically expressed at low endogenous levels, which explains the need for recombinant protein production for structural studies that require several hundreds of milligrams of purified protein (Andréll and Tate, 2013), and are intrinsically flexible due to the need to undergo conformational changes (Bill *et al*., 2011). Their hydrophobic nature reduces the compatibility of membrane proteins with aqueous buffers, and so solubilisation procedures that involve delipidation and replacement of the bilayer with detergent micelles have become a mainstay of membrane protein study (Bill *et al*., 2011). However, these are poor substitutes for membranes, as they lack lateral pressure which can be essential for maintaining the protein

fold and strip lipids from the structure that can be integral to the function and stability of the protein in question (Broecker *et al*., 2017).

However, in more recent years, a lot of work has been done to overcome these barriers and allow more comprehensive study of membrane proteins (Hardy *et al*., 2016), from improved overexpression systems and strategies (Andréll and Tate, 2013; Galluccio *et al*., 2022), to improving their stability and rigidity through mutagenesis (Harborne *et al*., 2020) and the advent of detergent-free alternatives for solubilisation (Guo, 2021), some of which allow retention of bound native lipids (Yeagle, 2014; Lee *et al*., 2016a) or relipidation (Cecchetti *et al*., 2021) and lipid-compatible techniques, such as native mass spectrometry (Corradi *et al*., 2019). Despite this, their innate and intimate association with the bilayer is challenging to conserve and the recent increase in understanding of the importance and role of lipids in protein study suggests the need for further work in this area (Laganowsky *et al*., 2014).

2.1.1 Lipid Interactions with Membrane Proteins

The lipid membrane, as well as providing the basis for compartmentalisation of cells and organelles, also acts as the site in which membrane-integral proteins are embedded, contributing to ~30% of the membrane surface area (Marrink *et al*., 2019). They are formed by two distinct leaflets of lipids, with their hydrophilic headgroups facing the solvent, and the hydrophobic tails between the two. Typically, the lipids are distributed asymmetrically between the leaflets, but this is not the only mechanism for non-uniform distribution of the bilayer contents, as they are known to disperse into nano-domains of lipids and membrane proteins (Marrink *et al*., 2019). The composition of the membrane is different across cell types and intracellular compartments, hinting at their importance in cellular physiology. Additionally, the lipid species and structures are different across kingdoms of life, contributing to over ~1000 lipid species (Corradi *et al*., 2019). This diversity leads to each membrane being able to have unique properties such as curvature (Huang *et al*., 2019), lateral pressure (Marsh, 2007), stiffness (Qi *et al*., 2015) and rate of lateral diffusion (Ridone *et al*., 2020) all of which impact the structure and function of the proteins inserted in the bilayer. The interplay between these properties and the associated proteins are termed general protein-lipid interactions, as they are dependent on the general physicochemical properties and composition of the bilayer, rather than distinct interactions with the protein itself.

In contrast, some of the effects of the bilayer on membrane proteins can be attributed to binding of specific lipid species to specific binding sites or through the protein modulating the bilayer surrounding it and forming a lipid “fingerprint” or “footprint” in the membrane. For example, the Na,K-ATPase structure is stabilised through interactions with POPS (Habeck *et al*., 2015), stimulated by neutral lipids, such as POPE, and are inhibited by POPC and

sphingomyelin (Haviv *et al*., 2013), through interactions at specific interactions sites. The bilayer is also able to facilitate conformational changes, as the tilting of the Ca^{2+}-ATPase throughout its catalytic cycle is governed by movement of phospholipids and anchoring of the protein in the bilayer (Norimatsu *et al*., 2017; Sweadner, 2017). Lipids have also been found to assist subunit oligomerisation of transporters, such as UapA (Pyle *et al*., 2018), LeuT, SemiSWEET and NhaA (Gupta *et al*., 2017). In addition to interactions between subunits to form a single functional unit, lipids have also been seen to control larger-scale organisation and clustering, such as of Band 3 anion exchanger 1 (Kalli and Reithmeier, 2018). However, there are only a limited number of techniques that can be used to study these interactions, including native mass-spectrometry (Gupta *et al*., 2017; Landreh *et al*., 2017; Pyle *et al*., 2018; Corradi *et al*., 2018), FRET (Loura *et al*., 2010; Dutta *et al*., 2014; Corey *et al*., 2018), click-chemistry (Gubbens and de Kroon, 2010; Peng *et al*., 2014; Corradi *et al*., 2019) and rarely through x-ray crystallography and cryo-EM (Yeagle, 2014).

This array of techniques is often supplemented by molecular dynamics (MD) simulations, which allow study of membrane proteins in native lipids. The simulation bilayer can more easily mimic the physicochemical membrane due to its continuous reach thanks to periodic boundary conditions and the presence of multiple atomistic representation of lipids available for modelling (Marrink *et al*., 2019; Corradi *et al*., 2019). Simulations have been used to study membrane proteins in lipid environments since 1994 (Woolf and Roux, 1994; Marrink *et al*., 2019), although this study of gramicidin A contained only 4390 atoms from the dimeric channel, 16 lipids and 649 water molecules and was simulated for 500 ps. For comparisons sake, current computational power has allowed entire viral envelopes to be simulated, with systems containing several millions of atoms (Huber *et al*., 2017).

As simulations are often able to link experimental outcomes to molecular detail and propose mechanisms for protein-lipid interactions that can be exploited in laboratory-based work, in this work I selected two clinically relevant membrane-integral proteins to better understand their function and structure in model bilayers. For diversity in lipid interactions and application of the findings of this work, one protein is found in all kingdoms of life excluding multicellular animals, while the other is found in almost all multicellular organisms. There is evidence that both of these proteins are regulated by lipids, and so through MD simulation I provided more clarity to these observations and hypotheses for further study.

2.2 Membrane-Integral Pyrophosphatases

Pyrophosphatases (PPases) are enzymes responsible for the reversible hydrolysis of the phosphoanhydride bond in pyrophosphate (PP_i) to two inorganic phosphate molecules (Kajander *et al*., 2013). PPases are subcategorised into three separate protein families: Family

I, Family II and membrane-integral pyrophosphatases (mPPases). Families I and II are evolutionarily unrelated soluble proteins. Despite their common enzymatic activity, mPPases are vastly different to the other PPase families. Firstly, the architecture and sequence of these protein families are unrelated. Ignoring the oligomeric structure, a single subunit of both Family I and mPPases consists of a single domain, while Family II PPases have two domains per subunit. mPPases are the only family that is embedded in the membrane: they have 15-17 transmembrane helices (TMH) per 70-81 kDa subunit (Lin *et al*., 2012; Kellosalo *et al*., 2012). Secondly, mPPases have only been reported as homodimers, whereas Family I PPases can form other oligomeric arrangements (Kankare *et al*., 1996; Lin *et al*., 2012; Kellosalo *et al*., 2012; Li *et al*., 2016; Tsai *et al*., 2019; Vidilaseris *et al*., 2019). Family I and II PPases are only responsible for removing excess waste PP_i from the cytoplasm, while mPPases are primary ion pumps: they couple hydrolysis to movement of H^+ and/or Na^+ across a membrane (Kajander *et al*., 2013) and thus generate a membrane potential, which contributes to a number of cellular functions, such as acidocalcisome and vacuole regulation and energisation (Shah *et al*., 2016). The catalytic activities of these protein families differ by orders of magnitude. The mPPases are the slowest as they can only hydrolyse ≈ 10 PP_i molecules per second, followed by Family I, which can hydrolyse ≈ 200 molecules per second, and Family II which can hydrolyse ≈ 2000 per second (Kajander *et al*., 2013).

mPPases are less abundant than soluble PPases but still occur in all kingdoms of life except fungi and multicellular animals (Kajander *et al*., 2013). In eukaryotes, they are localised in the membranes of organelles, such as the Golgi apparatus (Mitsuda *et al*., 2001), plant vacuoles (Gaxiola *et al*., 2007) or the acidocalcisomes of protists (Moreno and Docampo, 2009). They are also present in the inner cellular membrane of bacteria such as *Bacteroides vulgatus* (Luoto *et al*., 2013a). mPPases have been found to play a role in stress tolerance and plant maturation. Finally, mPPases promote energy efficiency and survival in numerous human pathogens, making them clinically relevant as potential drug targets (Luoto *et al*., 2011; Shah *et al*., 2016).

This evolutionarily ancient family evolved through a gene triplication; they thus consist of three structurally-conserved splayed 4-helix bundles made up of TMH 3-6, 9-12 and 13-16 (Au *et al*., 2006; Kajander *et al*., 2013) arranged with approximate three-fold symmetry perpendicular to the membrane plane. The bundles are structurally highly similar with RMSD/C_α values of 2.1-2.9 Å between them (Kellosalo *et al*., 2012) and also have 23.6-26.1% sequence identity (Kellosalo *et al*., 2012). Phylogenic analysis suggests that hydrolysis of PP_i to provide energy may have occurred prior to the adoption of ATP as the universal energy currency (Baltscheffsky *et al*., 1999). Therefore, mPPases may have been the first enzymes to couple phosphoanhydride bond formation/hydrolysis to chemiosmotic potential. Since the initial

discovery of mPPase in *Rhodospirillum rubrum*, seven different mPPase subfamilies have been discovered (**Table 2.1**) (Luoto *et al*., 2011; Tsai *et al*., 2014; Luoto *et al*., 2015). In brief, the different mPPases are subdivided into two main groups: (i) K^+ independent, which pump protons (H^+-PPases) and can be regulated by Na^+, and (ii) K^+ dependent, which can function in the absence of K^+, but require K^+ for maximal activity. Of these, there are H^+-PPases, Na^+ pumping (Na^+-PPase), and dual Na^+/H^+ pumping (Na^+/H^+-PPase) PPases (Kajander *et al*., 2013). Sequence analysis suggests that the first mPPases were Na^+-PPases, and that H^+-PPases evolved from these four independent times (Baykov *et al*., 2013). Additionally, it is likely that the evolution of Na^+/H^+-PPases occurred separately to the H^+-PPases (Luoto *et al*., 2013b).

Cation pumping specificity	Monovalent cation dependence	Semi-conserved glutamate location	Regulation	Hypothesised sodium binding site	Ref
H^+	K^+	6.57	-	-	(Tsai *et al*., 2014)
		6.53	-	-	(Tsai *et al*., 2014)
		5.43	-	-	(Tsai *et al*., 2014)
	-	6.53	Na^+ and K^+	Inhibitory	(Luoto *et al*., 2015)
			-	-	Luoto *et al*., 2011)
Na^+	K^+ and Na^+	6.53	-	Activating and Inhibitory	(Tsai *et al*., 2014)
Na^+ and H^+	K^+ and Na^+	6.53	-	Activating	(Tsai *et al*., 2014)

Table 2.1 The mPPase Subfamilies

2.2.1 Function and Relevance

2.2.1.1 Plants

Under physiological conditions in plants, H^+-PPases are predominantly localised to the tonoplast membrane surrounding the vacuole (Segami *et al*., 2014) and make up 10% of its protein components (Asaoka *et al*., 2016). The vacuole possesses multiple functions that require the large-scale movement of molecules across its membrane. The required membrane potential is generated by the vacuolar ATPase complex (V-ATPase) in combination with H^+-PPases (Kriegel *et al*., 2015). There is some controversy over the delineation of the roles of these proton pumps, as a V-ATPase knock-out strain was able to properly maintain acidification of the vacuole and normal function (Krebs *et al*., 2010). However, a more recent study suggested that lack of V-ATPase could not be compensated for by increased mPPase activity (Kriegel *et al*., 2015). The general consensus is that the role of H^+-PPase in stress tolerance is to replace V-ATPase activity when ATP levels are low (Maeshima, 2000), but this does not fully address the roles of the proton pumps under normal conditions.

H^+-PPases are important in plant maturation (Li *et al*., 2005) because they remove PP_i from the cytoplasm (Ferjani *et al*., 2011; Asaoka *et al*., 2016). PP_i is the by-product of many different cellular processes, including the biosynthesis of protein, RNA and, importantly for plants, cellulose (Maeshima, 2000). Removing the excess PP_i following these reactions is critical for driving these processes. Additionally, PP_i has a modulatory role as a biochemical intermediate of a number of enzymes (Heinonen, 2001), so tight control of its cytoplasmic availability is essential for normal cellular function. mPPase over-expression in *Arabidopsis thaliana* resulted in increased cell division and hyperplasia of different organs, in particular the leaves. In contrast, knock-out mutants and RNA interference studies showed severely disrupted root and shoot development. Each of these were linked to increased or decreased trafficking of the phytohormone auxin, which is known to mediate organogenesis (Li *et al*., 2005), suggesting a role for H^+-PPases in auxin regulation.

This role in auxin regulation was further highlighted in studies of transgenic plants over-expressing H^+-PPase genes. Multiple studies have shown that increased polar auxin transport upon mPPase over-expression is closely related to improved root development under stress conditions (Li *et al*., 2005; Park *et al*., 2005; Zhang *et al*., 2011; Pasapula *et al*., 2011). This plays a role in drought resistance, as the larger root system provides enhanced water absorption (Zhang *et al*., 2011). In addition to the increased root biomass mechanism, the effect of H^+-PPase over-expression on vacuolar function improved tolerance of drought and salinity. The increased electrochemical gradient may drive uptake of ions into the vacuole, producing an increase in osmotic potential and stimulating water uptake (Park *et al*., 2005;

Zhao *et al*., 2006; Brini *et al*., 2006; Zhang *et al*., 2011; Pasapula *et al*., 2011). Further evidence of the potential of mPPases to improve crop tolerance to suboptimal conditions were the reports of increased chlorophyll content, photosynthesis, leaf water content and fibre yield, with decreased cell membrane damage in transgenic cotton plants, as compared to wild-type under low water and high salt conditions (Lv *et al*., 2009; Pasapula *et al*., 2011). Interestingly, these effects may not be predominantly due to the proton pumping activity of the vacuolar mPPases. One study saw a *Vr*-PPase mutant lacking proton pumping activity, but retaining hydrolysis, was sufficient to rescue the stunted knock-out phenotypes (Asaoka *et al*., 2016). In addition to this, an *A. thaliana* plant H^+-PPase knockout saw no heterotrophic growth following germination, but this phenotype was rescued by soluble PPase expression, suggesting that effective PP_i clearance is the primary function of H^+-PPases during post germinative growth in Planta (Ferjani *et al*., 2011). It is not clear why this is the case, as all plant cells express soluble Family I PPases at concentrations that should be sufficient to clear the pyrophosphate generated. An explanation could be that the soluble and mPPases function cooperatively (Segami *et al*., 2018). In this model, the H^+-PPase functions as the major cytosolic PP_i-hydrolysis enzyme and the soluble PPases contribute to preventing accumulation to toxic levels, which would explain how soluble PPase expression was able to somewhat compensate for mPPase loss in the aforementioned study (Ferjani *et al*., 2011).

2.2.1.2 Protozoan Pathogens

A number of major human diseases are caused by protozoan parasites, for example, malaria (*Plasmodium ssp.*), toxoplasmosis (*Toxoplasma gondii*), trypanosomiasis (*Trypanosome spp.*) and leishmaniasis (*Leishmania spp.*) (Shah *et al*., 2016). These diseases each have a high prevalence and risk of fatality (Büscher *et al*., 2017; World Health Organization. Global Malaria Programme., 2018) or association with other diseases. For example, toxoplasmosis has been suggested to be associated with a number of conditions, such as psychiatric, neurological and neoplastic disorders (Torgerson and Mastroiacovo, 2013; Flegr *et al*., 2014). In addition to this, several protozoan strains responsible for malaria and trypanosomiasis have emerged that are resistant to most of the current treatment regimes (Büscher *et al*., 2017; World Health Organization. Global Malaria Programme., 2018). Therefore, there is a demand for novel therapeutics for these tropical diseases.

The protozoan parasite life cycle typically involves transitions between vectors and hosts and intracellular to extracellular environments, which means the protozoan cell must survive and adapt to several different conditions (Crompton *et al*., 2014). In terms of mPPase function, the most relevant change to overcome is the difference in osmotic pressure the cells experience in these different environments. The main protozoan mechanism for adjusting internal osmotic pressure involves the acidocalcisome (Docampo *et al*., 2013), where mPPases are localised

in protozoa (Scott *et al*., 1998; Marchesini *et al*., 2000). This is a small acidic compartment where numerous ions are stored, including polyphosphate, which the parasite hydrolyses or synthesises in response to osmotic stress and to release energy (Ruiz *et al*., 2001). The low pH of the acidocalcisome is crucial for its function, as loss of acidity can lead to a 10-fold decrease in stored polyphosphate levels (Lemercier *et al*., 2002), resulting in reduced capability to respond to osmotic changes. In addition to the effects on polyphosphate storage, there are detrimental effects on intracellular pH regulation, growth rate and final cell density (Lemercier *et al*., 2002), suggesting that the loss of mPPases has more widespread effects than just reduced osmotic regulation.

As mPPase function in protozoa is highly important, mPPases are a validated target for pharmacological intervention. Knock-down and knock-out studies in *Trypanosoma brucei* (Lemercier *et al*., 2002) *T. gondii* (Liu *et al*., 2014) and *P. falciparum* (Zhang *et al*., 2018) have demonstrated that mPPases are required for maintaining acidocalcisome acidification, parasitic virulence and *in vitro* asexual blood stage growth. Additionally, mPPase-inhibiting bisphosphonate derivatives retarded intracellular proliferation of *T. gondii* with no effect on host cells (Rodrigues *et al*., 2000).

2.2.1.3 Bacterial Pathogens

The most relevant mPPase-expressing bacterial genus to human health are *Bacteroides spp.*, especially *B. vulgatus* and *Bacteroides fragilis* (Luoto *et al*., 2013a), which form a mutualistic relationship with healthy individuals as part of the gastrointestinal microflora (Wexler, 2007). However, outside this environment, they can cause bacteraemia (presence of bacteria in the blood), intra-abdominal sepsis, appendicitis, gynaecological and skin infections, endocarditis, septic arthritis, and abscesses in tissues including the brain and female urogenital tract (Wexler, 2007). This is a major threat to human health, as *Bacteroides* reportedly have the highest resistance rates of all anaerobic pathogens (Wexler, 2007), and an associated mortality rate of over 19%, rising to 60% in untreated cases (Goldstein, 1996).

Bacterial mPPases are primarily found in species that exist under conditions of low-energy stress, such as the obligate anaerobes *Bacteroides* and deep-sea organisms (*Thermotoga maritima* and *Pyrobaculum aerophilum*) (Luoto *et al*., 2011). In these situations, mPPases may be essential to increasing energy efficiency and promoting survival. mPPases have also been shown to confer greater resistance to other stress conditions. One study revealed that expression of transgenic plant mPPases in *Escherichia coli* and *Saccharomyces cerevisiae*

Species	Ligand	Ions	Conformation	Mutation	PDB	Resolution (Å)	Noteworthy	Ref
V. radiata	IDP	5 Mg^{2+} K^{+}	Substrate-bound	-	4A01	2.35	First mPPase structure	(Lin *et al*., 2012)
	PO_4	2 Mg^{2+}	Relaxed product-bound	-	5GPJ	3.5		(Li *et al*., 2016)
	2 PO_4	5 Mg^{2+} K^{+}	Product-bound	-	6AFS	2.3	Demonstrated exit channel width affects function	(Tsai *et al*., 2019)
				$E^{6.57}Q$	6AFT	2.5		
				$L^{12.64}M$	6AFU	2.8		
				$L^{12.6}4K$	6AFV	2.7		
				$T^{5.36}D$	6AFW	2.2		
				$E^{5.33}A$	6AFX	2.3		
				$E^{5.33}S$	6AFY	2.4		
				$E^{5.33}H$	6AFZ	2.5		
T. maritima	-	Mg^{2+} Ca^{2+}	Resting	-	4AV3	2.6		(Kellosalo *et al*., 2012)
	2 PO_4	4 Mg^{2+} K^{+}	Product-bound	-	4AV6	4.0		
	IDP	5 Mg^{2+} Na^{+}	IDP-bound	-	5LZQ	3.5		(Li *et al*., 2016)
	WO_4	2 Mg^{2+}	Relaxed product-bound	-	5LZR	4.0		
	2 ATC IDP	5 Mg^{2+} Na^{+}	Locked	-	6QXA	3.7	Asymmetric structure with allosteric inhibitor	(Vidilaseris *et al*., 2019)

Table 2.2 The Experimentally Determined mPPase Structures

resulted in improved tolerance to heat, hydrogen peroxide and high salinity (Yoon *et al*., 2013). Therefore, I speculate that inhibiting mPPase activity could have effects on bacterial viability in response to stressors, similar to the effects seen in protozoan parasites and plants.

2.2.2 Structure and Mechanism

2.2.2.1 Structural Overview

In the current structures (**Table 2.2**), each subunit of the homodimeric mPPase is formed by 16 TMH, which form two concentric rings surrounding the four catalytic regions: the hydrolytic centre, the coupling funnel, the ion gate and the exit channel (Lin *et al*., 2012; Kellosalo *et al*., 2012; Li *et al*., 2016) (**Fig. 2.1**). The protein subunit-subunit interface is maintained by hydrophobic interactions and hydrogen bonds between residues on TMH 10 (I/R$^{10.33}$, N$^{10.42}$, S$^{10.45}$, K$^{10.49}$), 13 (L$^{13.19}$, L/V$^{13.21}$, N$^{13.22}$, M/V$^{13.23}$, I$^{13.33}$, Y$^{13.40}$, S$^{13.48}$, G$^{13.54}$, E$^{13.59}$,

TMH	Signifier	% Sequence identity	Residue in *Tm*-PPase	Residue in *Vr*-PPase	Residue in *Pa*-PPase	Residue in *Cp*-PPase
1	1.50	38.1	F17	F25	Y20	F17
2	2.50	27.6	K55	K94	R58	T55
3	3.50	99.5	S87	S153	S96	S92
4	4.50	92.9	G130	G194	G138	G133
5	5.50	86.4	R191	R242	R182	R190
5-6 Border	5.77	99.9	D218	D269	D210	D217
5-6 Border	6.26	96.2	D219	D270	D211	D218
6	6.50	99.7	D243	D294	D235	D242
7	7.50	46.3	G297	G334	A271	G278
8	8.50	75.8	L321	L361	L296	L304
9	9.50	91.1	G369	G411	G333	G341
10	10.50	84.3	S416	S458	S380	S388
11	11.50	98.1	D458	D500	D439	D438
12	12.50	99.7	K499	K541	K480	K479
13	13.50	99.4	V566	V597	V554	V541
14	14.50	79	M611	M642	F599	M586
15	15.50	94	A649	A680	S637	A624
16	16.50	94.5	K707	K742	K691	K680

Table 2.3 The Ballesteros and Weinstein Numbering System for mPPases

$R^{13.62}$) and 15 ($S^{15.45}$, $Q^{15.48}$) (Shah *et al*., 2016). Here, I use the Ballesteros and Weinstein numbering system, first introduced for GPCRs, so that all functional residues have the same index (Residuehelixnumber.offset) (Ballesteros and Weinstein, 1995), and which was adopted for mPPases (Tsai *et al*., 2014). In this system, each residue is assigned two numbers, one indicating the TMH and the other to describe its offset from a well-conserved residue close to the middle of the TMH (assigned position 50), as shown in **Table 2.3**. For example, K415 from *Tm*-PPase, is on TMH10 and is one before the conserved S416 and is thus designated $K^{10.49}$.

The inner ring is composed of six TMH (5, 6, 11, 12, 15, 16) and the outer ring is composed of the remaining 10 (Lin *et al*., 2012; Kellosalo *et al*., 2012; Li *et al*., 2016). A large hydrophilic region extends about 20 Å into the cytoplasm. This region has a hydrolytic centre, which is lined with highly conserved negative ($D^{5.61}$, $D^{5.65}$, $E^{5.76}$, $D^{5.77}$, $D^{6.35}$, $D^{6.39}$, $D^{6.43}$, $D^{11.57}$, $D^{15.61}$, $D^{16.31}$, $D^{16.35}$, $D^{16.39}$), positive ($K^{5.58}$, $K^{15.64}$, $K^{16.38}$) and polar ($N^{12.53}$) residues that are involved in substrate binding through both direct interactions and via coordination of Mg^{2+} and water. The hydrolytic centre is followed by the coupling funnel, which is comprised of an ionic network between TMH 5-6 ($R^{5.50}$, $K^{5.58}$, $D^{6.39}$, $D^{6.50}$), 11-12 ($D^{11.50}$, $K^{12.50}$) and 16 ($K^{16.38}$, $D^{16.39}$). Below this is the ionic gate, which is situated in the centre of the membrane and where the cation is bound prior to pumping. This region is made up of charged and hydrophilic residues ($S^{5.43}$, $D^{6.50}$, $E/S^{6.53}$, $S^{6.54}$, $E^{6.57}$, $D^{16.46}$, $K^{16.50}$). The hydrophobic gate located near the ionic gate and is formed by semi-conserved non-polar residues on TMH 5, 6, 12 and 16. Mutations of residues near the ionic and hydrophobic gates ($I^{12.54}$, $L^{12.64}$, $D^{16.46}$, $V^{16.54}$, $L^{16.57}$) to alanine decouple the enzyme so that mPPase can hydrolyse PP_i but does not pump, suggesting these regions are also essential for the coupling of hydrolysis to pumping (Asaoka *et al*., 2014). Finally, the residues of the exit channel are not highly conserved between mPPases. However, allosteric inhibitor binding to this region in one *Tm*-PPase subunit is nonetheless associated with a “locked” conformation (Vidilaseris *et al*., 2019) and exit channel width has been implicated in coupling hydrolysis and pumping (Tsai *et al*., 2019), thereby illustrating the importance of conformational changes in this region during the catalytic cycle.

2.2.2.2 Cation Specificity

Mutagenesis and structural studies showed that the residues $E^{6.53/6.57}$, $S^{6.54}$ and $D/N^{16.46}$ are involved in cation binding and pumping specificity (Asaoka *et al*., 2014; Li *et al*., 2016). The main residue regulating cation specificity is the semi-conserved glutamate on TMH6. Movement of this glutamate down one helical turn from $E^{6.53}$ to $E^{6.57}$ converts a Na^+-PPase to a H^+-PPase by destroying the Na^+ binding site in the substrate bound state (Li *et al*., 2016).

However, the reverse mutation does not yield a Na^+-PPase, indicating that the factors dictating cation specificity are more complex than solely the location of a single residue.

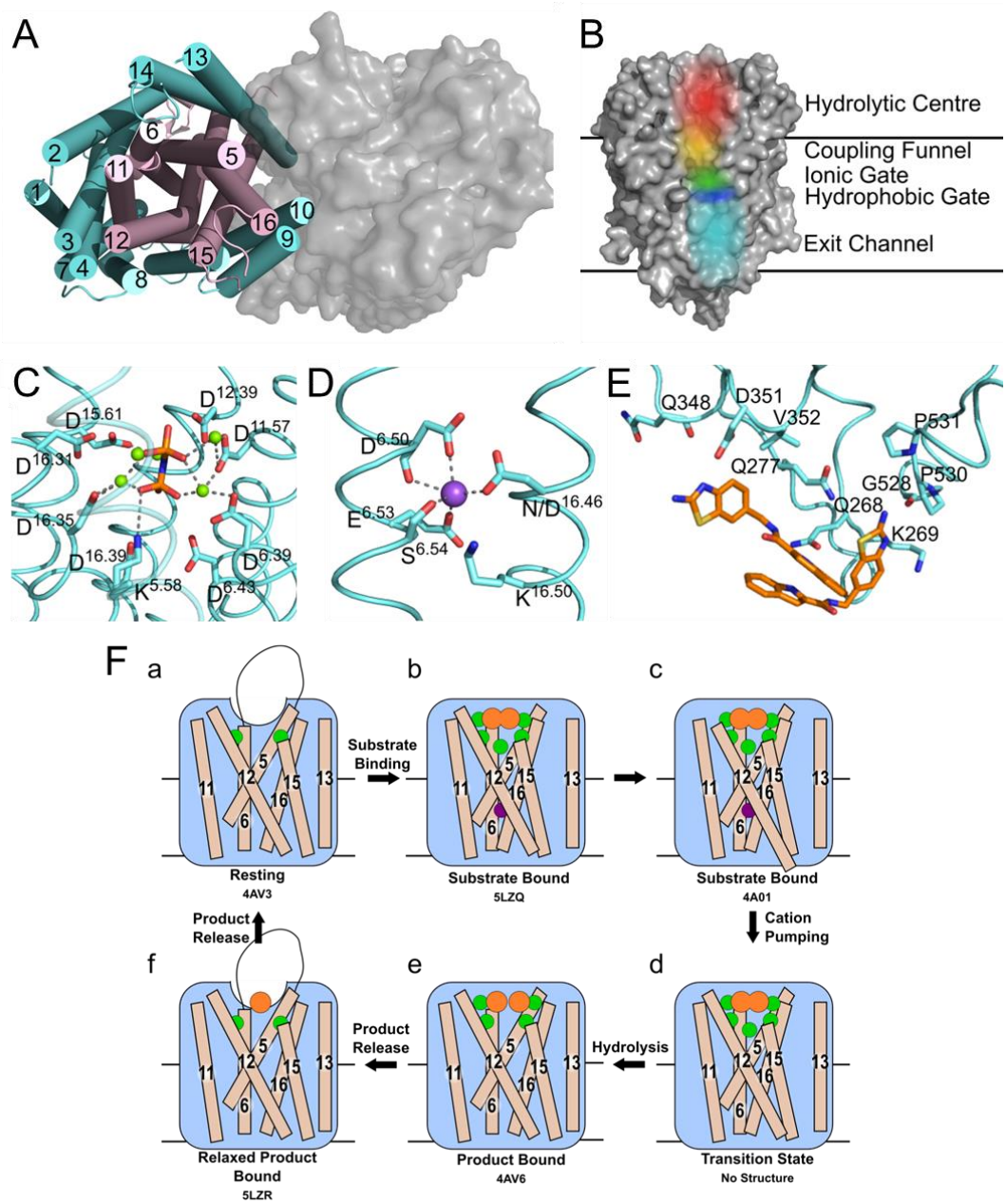


Figure 2.1 The Structural Arrangement and Catalytic Cycle of mPPases
A) the *Tm*-PPase resting state crystal structure (PDB: 4AV3) viewed from above to demonstrate the concentric circles of TMH of each subunit (outer ring in cyan and inner in pink) and the TMH contributing to the interface with the adjoining subunit (grey surface representation). **B)** A surface view of one mPPase subunit with each of the catalytic regions coloured (red, hydrolytic centre; yellow, coupling funnel; green, ionic gate; dark blue, hydrophobic gate; light blue, exit channel). The coordination of **C)** IDP and Mg^{2+} (green spheres) in the hydrolytic centre and **D)** of Na^+ (purple sphere) at the ionic gate in the IDP-bound *Tm*-PPase structure (PBD: 5LZQ). **E)** The binding of the two ATC molecules (orange sticks) to the β1-2 loop, loop 8-9, and loop 12-13, with the interacting residues shown at sticks (*Tm*-PPase numbering) (PDB: 6QXA). **F)** The "binding change" mechanism. In the resting state (a), a cation may be present, but its position has not been revealed structurally. When substrate binds, the hydrolytic centre is occluded by movement of the 5-6 loop and of TMH 11-12 and 15-16, and a cation localizes to the ionic gate (b). The cation is then pumped (c) and TMH 12 resumes its original position (d). Hydrolysis then occurs (e). Prior to the sequential release of the product P_i, the 5-6 loop, TMH 11-12 and 15-16 relax to their resting positions (f).

2.2.2.3 K+ Dependence and Independence

Mutation of $A^{12.46}$ to lysine in the active site converts a K^+-dependent enzyme to an independent enzyme (Belogurov and Lahti, 2002). Modelling this mutation suggested that the NH_3^+ group of the lysine sidechain functionally replaces the K^+ ion at physiological PP_i concentrations (Lin *et al*., 2012; Kellosalo *et al*., 2012). A systematic analysis of the potassium ion binding site or active site lysine ($K^+/K^{12.46}$ catalytic centre) across all mPPase subfamilies revealed that the lysine contributed to, but was not essential, for the activity of K^+ independent mPPases, as the activity was restored at high K^+ concentrations when mutated (Artukka *et al*., 2018). This suggests that $K^{12.46}$ masks a K^+ binding site, but there appear to be larger-scale differences between the two classes than just the identity of $A/K^{12.46}$ in the active site (See Inter-Subunit Communication).

2.2.2.4 Na^+ Regulatory Sites

The difference between Na^+- or H^+-PPases and Na^+/H^+-PPases was unclear prior to the discovery of Na^+ and K^+ regulated H^+-PPases (Luoto *et al*., 2015). These are a subfamily of K^+ independent H^+-PPases, in which low K^+ concentrations enhance H^+ transport but higher Na^+ and K^+ concentrations inhibit ion pumping and hydrolysis. This led to the hypothesis that there are two possible Na^+ binding sites; one in the ion conductance channel associated with activation of the mPPase, and the other located away from the conductance channel conferring inhibition of H^+ conductance, to which K^+ can bind with low affinity. This fits with the reports of H^+ translocation by Na^+-PPases in sub-physiological sodium concentrations (Nordbo *et al*., 2016), suggesting that loss of the second inhibitory site led to the evolution of the Na^+/H^+-PPases.

2.2.2.5 Mechanism of Pumping and Hydrolysis

Three mechanisms have been proposed for mPPase pumping and hydrolysis: (i) pumping occurs upon binding of substrate and prior to hydrolysis in a “binding change” type mechanism (Kellosalo *et al*., 2012) (**Fig. 2.1F**), (ii) the proton released upon hydrolysis triggers release of the ions at the ion gate *via* a Grotthus type mechanism (Lin *et al*., 2012), and (iii) the hydrolysis-generated proton is directly transported in a “direct-coupling” mechanism (Baykov *et al*., 2013). This first mechanism assumes that the nucleophilic proton may not be involved in pumping, whereas the second and third propose that the hydrolysis-generated proton is the one pumped, thereby assuming hydrolysis precedes cation-pumping. However, neither the second nor the third mechanisms account for Na^+ transport (Kajander *et al*., 2013), except for potentially *via* a “billiard-type” model in the third, where the Na^+ is pushed into the exit channel by the nucleophilic water proton (Baykov *et al*., 2013). In my opinion, models in which hydrolysis precedes cation pumping do not conform with the studies of non-hydrolysable

substrate analogues (IDP and methylene diphosphonate (MEDP)) binding to *Vr*-PPase (Li *et al*., 2016; Shah *et al*., 2017) resulting in a single ion turnover event.

The “binding change” mechanism unifies both sodium and proton pumping, and is supported by studies indicating that charge transfer – and so presumably ion pumping – does not require hydrolysis (Li *et al*., 2016; Shah *et al*., 2017) and so presumably precedes it in the full catalytic cycle. In this model, the substrate binds to the hydrolytic centre, which is then closed to the cytoplasm by ordering of the 5-6 loop and movement of TMHs 11-12 and 15-16 towards the centre of the coupling funnel. During these conformational changes, a cation is pumped out from the ionic gate. The increase in overall negative charge in this region leads to downwards movement of TMH 12 by 2 Å and an associated bend of TMH11, as observed structurally (Lin *et al*., 2012; Kellosalo *et al*., 2012; Li *et al*., 2016). This causes deprotonation of the aspartate pair $D^{6.43}$ and $D^{16.39}$, which activates the water nucleophile leading to hydrolysis (Kajander *et al*., 2013). This process has been demonstrated to be independent of cation pumping in exit channel mutants (Asaoka *et al*., 2014). These mutations were initially proposed to alter the position and angle of their helices and thereby uncouple the reactions (Asaoka *et al*., 2014). However, more recent structural characterisation of exit channel and hydrophobic gate mutants indicate that this uncoupling appear to be due to widening of the exit channel (Tsai *et al*., 2019), potentially allowing ion back-flow.

2.2.3 Asymmetry and Inter-Subunit Communication

The events occurring in each mPPase subunit during pumping and hydrolysis are relatively well understood (see above), but recent studies have provided evidence of asymmetry and subunit interdependence. Firstly, the role of the $K^+/K^{12.46}$ catalytic centre in inter-subunit communication, as inferred from the effects of excess substrate was published (Artukka *et al*., 2018) (See Inter-Subunit Communication). Shortly after this, the first structural evidence of asymmetry, further functional evidence and a putative mechanism involving both subunits were reported (Vidilaseris *et al*., 2019). The earliest reports of subunit interdependence were in the 1990s in a series of studies utilising radiation-induced damage to identify the functional unit size of mPPases (Wu *et al*., 1991; Sarafian *et al*., 1992; Tzeng *et al*., 1996). Several of these showed that one impaired subunit conferred compromised function to the unaffected subunit (See Inter-Subunit Communication).

2.2.3.1 Functional Asymmetry

Vidilaseris and colleagues (2019) reported that binding of the first substrate increased the affinity of the second subunit to PP_i. They also observed that the binding of first substrate molecule potentiated binding of the second. This suggests positive co-operativity for substrate binding, and thus they proposed the following mechanism: the first substrate binds to the first

subunit leading to cation translocation in this subunit, this causes conformational changes in the second subunit to optimize PP_i binding. Following the binding of the second substrate molecule, hydrolysis occurs in the first subunit and ion pumping in the second. This then allows hydrolysis in the second subunit and product release. However, the same study (Vidilaseris *et al*., 2019) also reported a 20-fold decrease in the maximal rate of hydrolysis when both subunits bind substrate, which does not appear to fit with their proposed asymmetric mechanism.

In contrast, Artukka and colleagues (2018) reported a decrease in affinity for PP_i at the second subunit following substrate binding to the first. The binding of fluorescein 5'-isothiocyanate (FITC) to a lysine believed to be in the catalytic centre ($K^{12.50}$) also decreases the affinity for FITC in the second subunit (Yang *et al*., 2000), providing further support for an allosteric mechanism. Binding at the first site thus presumably distorts the active site in the second subunit, reducing its affinity for PP_i/FITC. Negative co-operativity for substrate binding is highly relevant, because excess substrate inhibits hydrolysis in all wild-type mesophilic mPPases studied so far (Artukka *et al*., 2018), so excess substrate may play a role in proton pumping inhibition (Luoto *et al*., 2015). As there is no evidence of an alternative PP_i binding site, and the hydrolytic centre and coupling funnel could not accommodate further PP_i molecules, the inhibition must be a result of substrate binding to the second subunit. This would only be possible in excess substrate concentrations, due to the decreased affinity at the second site. Similar to Vidilaseris and colleagues (2019), this study also saw a 3 to 16-fold decrease in maximal velocity of hydrolysis in both subunits upon binding of the second substrate. This again supports the model in which excess substrate inhibits hydrolysis through binding at this second site, thus suggesting that at moderate PP_i concentrations only one subunit operates at any given time. The direct implication of this is that the symmetrical IDP bound structures are potentially not mechanistically relevant at typical PP_i concentrations, but a possible artefact of the 4 mM minimum IDP concentration during crystallization (Lin *et al*., 2012; Li *et al*., 2016). These papers by Artukka and colleagues (2018) and Vidilaseris and colleagues (2019) are the first two definitive studies of possible asymmetry: further data are needed to understand fully the mechanism involving both subunits.

What could explain the inconsistencies between these two studies? Firstly, one study investigated thermophilic protein in an assay at 71 °C (Vidilaseris *et al*., 2019), whereas, the other used mesophilic protein at 30 °C (Artukka *et al*., 2018). The behaviour could indicate different properties between thermo- and mesophilic proteins, or different activities at different temperatures. Secondly, one study was performed with purified protein in detergent micelles (Vidilaseris *et al*., 2019) and the other used unpurified protein in inner membrane vesicles (Artukka *et al*., 2018). This could suggest that the lipid environment of the enzyme could be

involved in regulating the dimeric interface, a concept for which there is increasing evidence (Gupta *et al*., 2017; Pyle *et al*., 2018).

2.2.3.2 Structural Asymmetry

Kinetic and structural studies showed that the non-phosphorous allosteric inhibitor N-[(2-amino-6-benzothiazolyl)methyl]-1H-indole-2-carboxamide (ATC) binds as a dimer to one of the subunits in *Tm*-PPase (Vidilaseris *et al*., 2019) thereby trapping an asymmetric structure, with RMSD/C_α values reaching 1.6 Å for some of the loops versus a highest RMSD/C_α of 0.5 Å for the equivalent loops in the symmetrical IDP-bound structure (PDB: 5LZQ). The ATC dimer bound via the β1-2 loop (Q268, K269, Q277), loop 8-9 (Q348, D351, V352) and loop 12-13 (G528, P530, P531) (*Tm*-PPase numbering). The chief difference between the 5LZQ and 6QXA structures is the presence of ATC in the crystals. Due to the ATC, the β1-2 loop moves closer to loop 12-13 and away from the dimer interface to an angle similar to that seen in the resting and product bound structures, indicating closure of the exit channel. Loop 12-13 also experienced some slight movements, but TMH12 still moves downwards by about 2 Å compared to the resting state. In addition, the coordination of Na^+ at the ionic gate changes from pentacoordination to tetracoordination and the ion is displaced by about 1.2 Å from the IDP-only structure. Thus, the ATC-bound subunit appears to be trapped in a closed exit channel state and cannot perform hydrolysis, and the unbound subunit cannot undergo its catalytic cycle, due to the ATC-bound subunit restricting its motions.

2.2.3.3 Inter-Subunit Communication

Artukka and colleagues (2018) demonstrated that the wild-type $K^+/K^{12.46}$ centre was important for inter-subunit communication. Substrate inhibition only occurred in K^+ independent mPPases when the native $K^{12.46}$ residue was present. In the $K^{12.46}$A substituted enzymes there was no inhibition. Similarly, K^+ dependent mPPases were not inhibited by excess PP_i unless K^+ was present, even in $A^{12.46}$K substituted enzymes. Consequently, this suggests the lysine and potassium ion are unable to perfectly replace each other, despite the findings of previous modelling studies (Kellosalo *et al*., 2012). This is further evidence that inter-subunit communication in response to excess PP_i is modulated by the native $K^+/K^{12.46}$ centre, as its presence is required for excess substrate inhibition.

The recent asymmetrical ATC structure (see above) demonstrated direct linkage of TMH 12 to the dimer interface (Vidilaseris *et al*., 2019). Vidilaseris and colleagues suggest that, in the presence of K^+, the movement of TMH 12 upon substrate binding to one subunit induces conformational changes in the dimer interface, which are then propagated to the other subunit. Moreover, when residues close to $A^{12.46}$ were mutated to alanine, proton pumping and hydrolysis were uncoupled (Asaoka *et al*., 2014). This may be because this region is

implicated in inter-subunit communication for cation translocation at normal PP_i levels, and thus when it is mutated, cation pumping ceased. This is further supported by the fact that mPPases are dimers, even though all the required catalytic domains are contained within a single monomer.

Early studies of the functional unit size of mPPases demonstrated that mPPases had to be functional dimers for cation pumping (Sarafian *et al*., 1992) but a single catalytic subunit was sufficient for hydrolysis (Sarafian *et al*., 1992; Tzeng *et al*., 1996). Although these studies were limited by the utilization of radiation-induced destruction, a technique which damages the entire molecule, not just specific sites (Kempner, 1993), they are still relevant when interpreting more recent evidence of subunit interdependence.

2.3 The Mechanosensitive Ion Channel PIEZO1

The PIEZO protein family consists of two members: PIEZO1 and PIEZO2. These represent a distinct class of mammalian ion channel as they work to convert mechanical stimuli to ion transduction and subsequent downstream processes. The evidence that these channels existed was first published by Corey and Hudspeth (1979), but the specific channels were not identified until 2010 (Coste *et al*., 2010). These channels possess a unique protein fold that causes perturbation of the surrounding membrane, which may explain their mechanosensing properties (Guo and MacKinnon, 2017). Mechanically activated ion channels are an important component of the function and survival of organisms, and the most well studied of these is the large-conductance mechanosensitive channel (MscL) family in bacteria, which protect the cell from osmotic shock (Martinac *et al*., 1987). However, MscL and PIEZO channels differ in a number of important ways. MscL opening is characterised by a large pore area which does not allow ion selectivity but encourages large conductance of 3 nS (Sukharev *et al*., 1993), whereas PIEZO channels are cation selective, which requires a narrow pore, and have lower conductance values of 30 pS (Coste *et al*., 2010). Therefore, the distinct properties, sequence and domains of the PIEZO channels cells could not be based on insights from the MscL proteins or even other mammalian ion channels that have been implicated in mechanosensation (Delmas *et al*., 2022).

These proteins responsible for mechanosensation were identified by Coste and colleagues (2010), leading to the 2021 Nobel Prize, through a knockdown screen that demonstrated loss of mechanically activated currents, but prior to their discovery many processes were attributed to them. Following their discovery, a vast amount of work has been performed to understand their structure, function and modulation. Much of this work has focused on the PIEZO channels found in *Mus musculus*, but the channel is found in almost all multicellular and unicellular eukaryotic organisms, with the exception of yeasts (Coste *et al*., 2010). This work has led to

the identification of PIEZO channels as potential targets for pharmacological intervention (Xiao, 2020; Tang *et al*., 2022), as they are involved in processes from the senses to embryogenesis (Delmas *et al*., 2022), and point mutations in the channel can lead to disease (Song *et al*., 2022a).

The two members of the PIEZO channel family, are delineated in their expression profiles and roles, as PIEZO2 is implicated in mechanical stimuli that take a tactile or auditory form (Wang *et al*., 2019), whereas PIEZO1 is strongly linked to systems that experience mechanical stimuli in the form of flow, shear stress, membrane tension and cell swelling (Delmas *et al*., 2022). There are differences between the functional properties of the two channels beyond expression (Lewis and Grandl, 2020), but this work focuses on the PIEZO1 channel due to its more extensive structural and simulation based characterisation.

2.3.1 Function and relevance

2.3.1.1 Physiological PIEZO1

Due to the expression of PIEZO1 throughout the body (Coste *et al*., 2010; Volkers *et al*., 2015; Delmas *et al*., 2022), it has been linked to several physiological processes, including somatosensation (Hill *et al*., 2022), urinary control (Dalghi *et al*., 2021), cardiovascular regulation (Li *et al*., 2014; Cahalan *et al*., 2015) with impacts on physical activity (Bartoli *et al*., 2022), lymphatic maintenance (Nonomura *et al*., 2018), wound healing (Holt *et al*., 2021) and immune system processes (Solis *et al*., 2019). The apparent near ubiquitous nature of PIEZO1's activity and its global knockout being incompatible with life in mice, demonstrates its ability to be activated by a range of mechanical stimuli and importance as a mechanosensor on many different cell types (Delmas *et al*., 2022).

Alongside forming trimeric oligomers with other PIEZO channel subunits (Guo and MacKinnon, 2017), these proteins have been found to work cooperatively in large scale clusters of PIEZO1 channels (Ridone *et al*., 2020) and interact with other membrane proteins (Volkers *et al*., 2015). The interaction with these membrane proteins has the ability to modulate function, for example polycystin 2 is able to inhibit PIEZO1 function (Peyronnet *et al*., 2013), whereas stomatin-like protein 3 is required for PIEZO channel function in touch neurons (Volkers *et al*., 2015). The inter-play between PIEZO1 and other proteins indicates that it has great tunability, and so is able to take on multiple different roles in different cell types as needed.

2.3.1.2 Pathophysiological PIEZO1

As PIEZO1 appears to be so widely expressed and essential for many physiological processes, where PIEZO2 or other mechanosensitive mechanisms are unable to compensate for its loss (Delmas *et al*., 2022), it is easy to imagine how mutation to this channel can impact

its function and lead to disease (Song *et al*., 2022a). Currently, over 25 mutations in PIEZO1 have been linked to diseases, but due to global knockout being lethal in mice, it is likely that even more mutations exist but are incompatible with life. The primary disorders linked to PIEZO1 mutation are dehydrated hereditary stomatocytosis and congenital lymphatic dysplasia.

Dehydrated hereditary stomatocytosis linked to PIEZO1 mutation is characterised by dehydration of red blood cells and subsequent changes to their osmotic fragility (Song *et al*., 2022a). Therefore, haemolytic anaemia is induced in the patient, in which red blood cells are damaged due to their inability to regulate their volume (Jankovsky *et al*., 2021), whereas knock out study of PIEZO1 in red blood cells causes the inverse effect (Cahalan *et al*., 2015). This disease is caused by gain of function mutations to PIEZO1, including R2456H, R2488Q and E2496ELE (note that human numbering is used), which prolongs and increases the cation influx by slowing inactivation (Bae *et al*., 2013a; Evans *et al*., 2020). In contrast, congenital lymphatic dysplasia is linked to loss of function mutations in PIEZO1 (Lukacs *et al*., 2015), in which truncated or missense mutation attenuates channel function. It is thought that the mutated channel is unable to be effectively expressed at the cell surface, and is therefore unable to contribute to lymphatic development (Lukacs *et al*., 2015), rather than direct alteration of the channel function and properties of ion conductance. From the properties of the substituted amino acids themselves, some hypotheses for how mutation alters PIEZO1 activity can be formed, but structural information has been key in determining the role of different regions and characteristics of the protein in disease (Guo and MacKinnon, 2017).

2.3.2 Structure and Mechanism

2.3.2.1 Structural Overview

The structures of PIEZO channels have remained somewhat elusive. Until 2015, only the structure of the extracellular cap domain had been resolved (Kamajaya *et al*., 2014), but in the 8 subsequent years, 8 cryo-EM structures of PIEZO channels have been determined (**Table 2.4**) (Ge *et al*., 2015; Guo and MacKinnon, 2017; Saotome *et al*., 2018; Zhao *et al*., 2018; Wang *et al*., 2019; Geng *et al*., 2020; Yang *et al*., 2022). The majority of these structures have been of the PIEZO1 channel from *Mus musculus* (mPIEZO1), which have provided the insight that mPIEZO1 is a large, curved, trimeric membrane protein composed of a central pore, above which sits the extracellular cap domain, and from which triskelion blades project. The pore and extracellular cap domain are each made up of elements of three subunits, whereas the blades do not appear to directly interact with each other in the structures (**Fig 2.2**). The blade structure is made up of nine 4-helix helical bundle units (HBUs), that end with a final two TMH that line the pore (37 and 38).

Species	Conformation	PDB	Resolution (Å)	Noteworthy	Ref
Caenorhabditis elegans	-	4PKE	2.5	Structure of the extracellular cap domain only	(Kamajaya *et al*., 2014)
	-	4PKX	2.54	Structure with a disease-causing mutation	
Mus musculus	Curved	3JAC	4.8	First structure of the membrane soluble section of a PIEZO channel	(Ge *et al*., 2015)
	Curved	6B3R	3.7	Used for structural modelling	(Guo and MacKinnon, 2017)
	Curved	6BPZ	3.8	-	(Saotome *et al*., 2018)
	Curved	5Z10	3.97	-	(Zhao *et al*., 2018)
	Curved	6KG7	3.8	First structure of PIEZO2 First full-length blade structure	(Wang *et al*., 2019)
	Curved	6LQI	4.5	-	(Geng *et al*., 2020)
	Curved	7WLT	3.46	First structure of a PIEZO channel in a lipid bilayer	(Yang *et al*., 2022)
	Flattened	7WLU	6.81	First flattened structure of a PIEZO channel	

Table 2.4 The Experimentally Determined PIEZO Structures

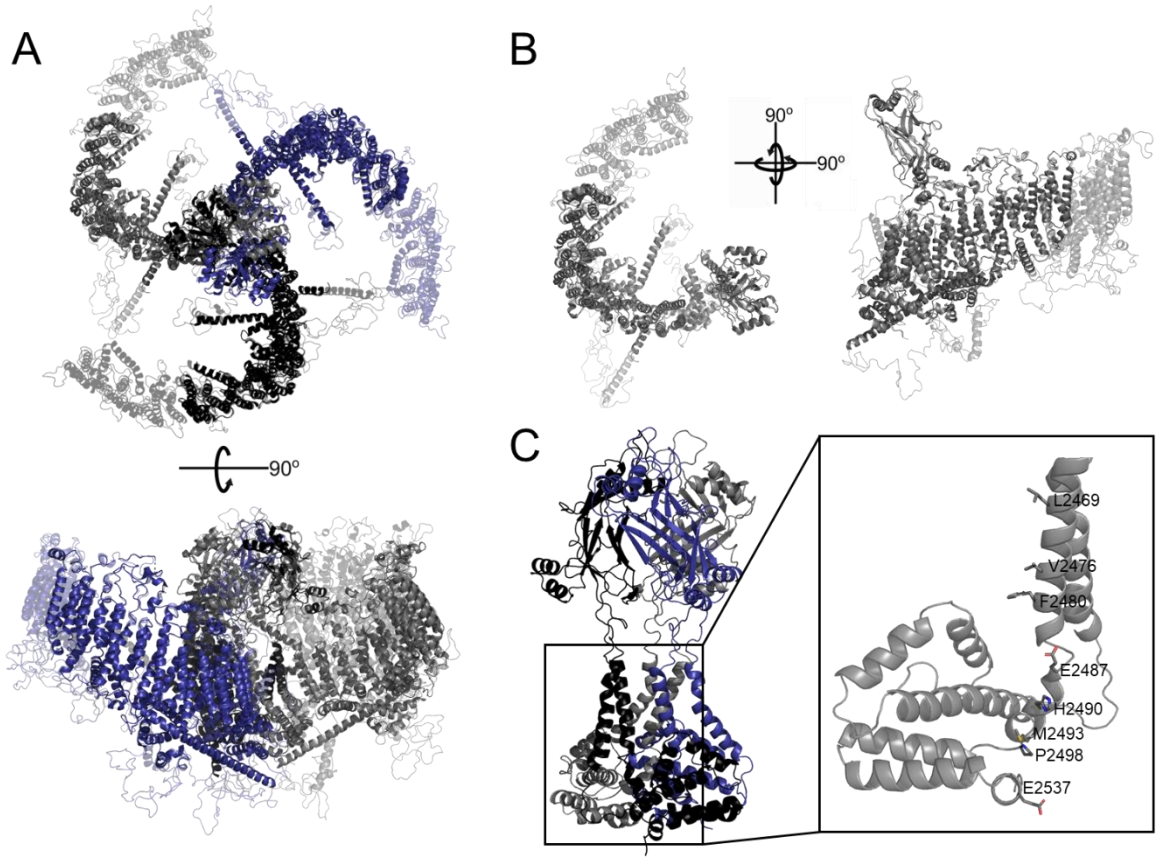


Figure 2.2 The Structural Arrangement of mPIEZO1
A) The structure of mPIEZO1 as determined by high-resolution cryo-EM studies superimposed on a full-length model of the protein (transparent helices), where each subunit is coloured in black, dark grey and dark blue, viewed from both the extracellular side and the membrane plane. **B)** The detail of each blade, which is composed of 38 transmembrane helices with adjoining membrane parallel amphipathic helices, an extracellular cap, where again the full-length model is shown in transparent to demonstrate missing regions. **C)** The pore of mPIEZO1 demonstrating the hydrophobic gate and pore lining helices. (PDB: 6B3R) and full-length model reported by (Chong *et al*., 2021).

A lot of this information was gained from the initial structure by Ge and colleagues (2015), but due to a number of errors in the structure, such as modelling two helices in the place of one and planar arrangement of the blades, much of the insight into the structural architecture and their biological meaning has been gained from the structure resolved by Guo and MacKinnon (2017). Despite the structural understanding gained from these mPIEZO1 structures, all of the available data are missing significant portions of the protein. In the case of the Guo and MacKinnon (2017) structure, only 1518 of 2547 residues are modelled, with the missing residues making up the first 576 residues of the blade, and several intracellular loops. Additionally, the resolution of a number of these structures are only verging on the level of data where backbone tracing and side chain allocation are possible (Arkhipova *et al*., 2017). Despite this, the structural insights into the resolved blade structure and pore have allowed comparison to other mechanosensing ion channels and a mechanism of stimuli propagation to the pore to be proposed. Additionally, the extracellular cap domain and intracellular loops have provided more ideas for the mechanism of sensing different mechanical stimuli.

2.3.2.2 Pore

The pore is one of the major sites at which the three protein subunits come together, as each of the final two helices of the blade, TMH37 and TMH38 domain swap between the subunits and form a central pore lined by the E2487, H2490, M2493, P2536 and E2537 residues (note that mPEIZO1 numbering is used) (Guo and MacKinnon, 2017). Structural studies suggest that the glutamate residues at either end of the ion permeation pathway contribute to the cation selectivity of the protein, and their mutation abolishes the cation selectivity of the channel (Zhao *et al*., 2016). As far as structural methods allow, the pore is undeniably closed in almost all of the mPIEZO1 structures available, only reaching a maximum diameter of 0.4 Å at the level of P2536 (Guo and MacKinnon, 2017), as experiments on the human PIEZO1 indicate that the pore should open to at least 4 Å diameter (Gnanasambandam *et al*., 2015).

The similarity in pore architecture between mPIEZO1 and other ion channels, especially P2X (Kawate *et al*., 2009), gave rise to the hypothesis that the ions enter the pore *via* lateral fenestrations beneath the extracellular cap domain and then traverse the ion conduction pathway (Kellenberger and Grutter, 2015; Guo and MacKinnon, 2017), mutagenesis of these fenestrations has added evidence to this hypothesis (Geng *et al*., 2020), but other mechanisms such as translocation through the extracellular cap domain have not been entirely excluded (Wang *et al*., 2019; Lewis and Grandl, 2020). Entrance through the lateral fenestrations would implicate the regions surrounding the pore in transduction of the mechanical stimuli to the pore, as these are not present in other structurally similar cation channels that are unable to sense force. These regions are the elbow and base, which directly connect the blade to TMH37, and the hairpin and pore extension C-terminal helix, which sit beneath the pore. Additionally, there has been interest in a plug-and-latch mechanism, but these have not been demonstrated structurally (Geng *et al*., 2020).

Despite the other structures appearing to be unequivocally closed, the flattened structure of mPIEZO1 does not provide a definitive pore state, as the resolution and electron density does not provide accurate side chain density (Yang *et al*., 2022), and so the measurements that may provide clarity as to its state are unreliable. Despite this, the authors report no major changes to M2493, P2536 or E2537, which might suggest that this state is inactive, implying more than membrane flattening is required for the pore to open. There are a number of suggested gating mechanisms which might explain this. For example, the structure of mPIEZO2 suggests the presence of a hydrophobic gate at residues L2469, V2476 and F2480 (Wang *et al*., 2019), MD and structural studies suggest the movement of a bound lipid at the protein pore (De Vecchis *et al*., 2021; Yang *et al*., 2022), the motion of a plug and latch domain (Geng *et al*., 2020), and the decoupling of the cap domain from the protein (Lewis and Grandl, 2020; Yang *et al*., 2022).

2.3.2.3 Extracellular Cap Domain

The extracellular cap domain sits above the channel pore and is formed by a novel β-sandwich fold from each of the protein subunits (Kamajaya *et al*., 2014; Ge *et al*., 2015), where three β-pleated sheets come together while being flanked by two α-helices. Despite focussing on a single soluble domain of PIEZO1 rather than the entire structure, which is likely inaccessible through x-ray crystallography, the insights into the arrangements of the extracellular cap domain and the impact of a disease-causing mutation on its structure have helped inform understanding of how this domain contributes to protein function.

Kamajaya and colleagues (2014) concluded that this domain was unlikely to participate in oligomerisation of the PIEZO1 channel, which has since been disproven by other structural studies demonstrating that the cap is made up of the three subunits and an extensive interface exists between the three (Ge *et al*., 2015; Guo and MacKinnon, 2017; Saotome *et al*., 2018; Zhao *et al*., 2018; Wang *et al*., 2019; Geng *et al*., 2020; Yang *et al*., 2022), but the overall structure of each monomeric extracellular cap has not been disputed. In addition to revealing a novel fold, this structural study examined the effect of the disease-causing mutation M2225R (human PIEZO1 numbering) linked to dehydrated hereditary stomatocytosis which is mapped to this region of the protein and results in delayed activation and slow inactivation (Bae *et al*., 2013a). Although this is a significant change in size and charge, it did not result in changes to the structure of this domain, although it was suggested that this change could have impacts on the interactions between the cap and other regions of the protein (Kamajaya *et al*., 2014). This is of particular interest, as just two point mutations, including M2225R, to this domain can abolish channel inactivation (Bae *et al*., 2013b). Therefore, it is clear that the extracellular cap domain is involved in protein function.

This led to the investigation of this domain to better understand its role in PIEZO1 activity. Mutation of the domain led to insensitivity to the role of pH on inactivation following mutation to the extracellular cap domain (Bae *et al*., 2015), further confirming its role. To understand this on a more granular level, chimerical substitution of regions of the mPIEZO1 cap with regions of the mPIEZO2 cap allowed the inactivation kinetics of mPIEZO2 to be transferred to mPIEZO1 (Lewis and Grandl, 2020), demonstrating that specific sequence and structural motifs are required for different channel kinetics. These functional data are supported by the rotation of the cap domain in the flattened mPIEZO1 structure, indicating that it may be essential for activation of the channel (Yang *et al*., 2022). However, structural studies of the curved protein have also demonstrated several classes of extracellular cap domain conformation, with typically a 15° lateral rotation of the cap defining each of these (Ge *et al*., 2015; Saotome *et al*., 2018; Zhao *et al*., 2019). Lewis and Grandl (2020) demonstrated the importance of motion of this cap by crosslinking it to the blade of mPIEZO1 and thereby

abolishing mechanical activation, with the suggestion that that ion translocation occurs through the extracellular cap, and so alteration of the arrangement of the cap domain may act as an inactivation gate composed of K2479, L2475 and V2476. However, this is not a fully accepted translocation pathway (Guo and MacKinnon, 2017), and so it remains unclear how the structure and motion of the cap domain is implicated in the mechanism of PIEZO activation and inactivation.

2.3.2.4 Blade Structure

The blade comprises the majority of the PIEZO protein, being composed of 36 helices arranged into discrete HBUs each made up of 4 TMH and a membrane parallel helix (MPH) that projects from the sequentially first TMH of the four (Ge *et al*., 2015; Guo and MacKinnon, 2017; Saotome *et al*., 2018; Zhao *et al*., 2018; Wang *et al*., 2019; Geng *et al*., 2020; Yang *et al*., 2022). Mutation of the extracellular loops of the blade HBUs results in poor expression and targeting to the membrane (Zhao *et al*., 2018), indicating their importance for protein production and trafficking. In addition to these HBUs and MPHs, the proximal three HBUs sit above a long helical beam, which has been linked to the force-sensing capabilities of PIEZO channels (Saotome *et al*., 2018; Zhao *et al*., 2018; Wang *et al*., 2019). Connecting the blade to the pore is a region composed of three helices, two of which form an elbow and the third a base that sits parallel to the membrane (Guo and MacKinnon, 2017). The lack of structurally resolved details of the full PIEZO1 blade structure has led to this domain being overlooked, but its structural details and dynamics are key to PIEZO1 function. The structure of the blades causes the HBUs to spiral outwards from the pore in the XY plane, but also in the Z direction, which indicates that they are able to form a membrane dome that projects into the cytoplasm, which has been verified through cryo-EM studies (Guo and MacKinnon, 2017; Yang *et al*., 2022) and MD simulations which emphasise the length of the blade in dome depth (Chong *et al*., 2021). The formation of this dome is assumed to be needed to allow the blades of PIEZO channels to sense mechanical stimuli and the pathway of stimuli propagation to the pore, and so understanding the blade structure is essential to forming and testing hypotheses involving mechanical stimuli.

The structure of the full-length blade of mPIEZO1 has not been structurally resolved, likely due to its high level of flexibility in both detergent micelles and liposomes (Ge *et al*., 2015; Guo and MacKinnon, 2017; Saotome *et al*., 2018; Zhao *et al*., 2018; Geng *et al*., 2020; Yang *et al*., 2022). This flexibility can be captured structurally, resulting in the identification of distinct blade conformations in the XY plane owing to lateral motion of the distal HBUs (Saotome *et al*., 2018), which has been suggested to be integral to mechanosensing (Saotome *et al*., 2018; De Vecchis *et al*., 2021). Despite the lack of structural information on the mPIEZO1, the structure of the blade from mPIEZO2 allows some insights into the distal section of this domain

(Wang *et al*., 2019), and models of the full-length channel have also been reported by Chong and colleagues (2021) and by AlphaFold2 (Jumper *et al*., 2021; Varadi *et al*., 2022).

The structure of the N-terminal end of the blade may appear to be trivial to model, as there is high sequence identity between the resolved bundles and the ones to be predicted, which should correspond to a high level of structural predictability (Lushington, 2014). However, it is not only 12 TMH missing from a planar structure that need to be modelled, but their spatial arrangement compared to the rest of the protein needs to be determined, as this will impact understanding of the formation and depth of the membrane dome and the impact of flattening on the structure. The modelled structures of the blade, both by Chong and colleagues (2021) and AlphaFold2 (Varadi *et al*., 2022) suggest that the blade continues to curve upwards in the Z direction, as well as in the XY plane. This is consistent with the structure of mPIEZO2 where the blade continues its trajectory (Wang *et al*., 2019). However, study of the mPIEZO1 structure in liposomes suggests that the distal region of the blade may actually adopt a planar structure, as this region is where the dome flattens in liposomes (Guo and MacKinnon, 2017; Yang *et al*., 2022). This may be explained by a difference in the structure, and therefore functional properties, of mPIEZO2 compared to mPIEZO1, or the result of studying the structure in detergent micelles rather than the native bilayer.

The flattening of the blade upon membrane tension and pressure was predicted prior to the structural study of flattened mPIEZO1, as it is a common mechanism of mechanosensitive channel activation (Ursell *et al*., 2008; Phillips *et al*., 2009; Pliotas *et al*., 2015). Expansion of the membrane area by flattening reduces the potential energy of the protein system, which can then be converted into activation and opening of the channel (Guo and MacKinnon, 2017). Experimental studies have also demonstrated flattening of the protein following the application of force, such as in HS-AFM by Lin and colleagues (2019), confirming that the blade appears to be the main site of mechanical sensation and conformational change upon application of pressure.

By exploiting the ability to mPIEZO1 in an outside-out configuration in small (10 nm radius) liposomes, as opposed to the more favourable outside-in arrangement (Guo and MacKinnon, 2017), Yang and colleagues (2022) were able to determine the structure of the flattened mPIEZO1 to understand the structural changes that occur in the presence of membrane tension. Interestingly, in the flattened structure the distal HBUs continued to be unresolved, indicating that even upon change in conformation these areas remain flexible. The blades were flat in Z but continued to curve in their propeller like shape away from the pore in the now planar membrane. This is somewhat corroborated by an MD study of lateral membrane

tension on mPIEZO1 opening (De Vecchis *et al*., 2021), in which the blade also appeared to straighten upon membrane tension and pore opening.

2.3.2.5 Mechanism of Force Sensing

The exact mechanism of force sensing by PIEZO1 is unclear, as two hypotheses exist: force from filaments (i.e., where tethers to extra- or intracellular components exert force on the protein and trigger its opening) and force from lipids (i.e., where lateral membrane tension and subsequent increase in area per lipid triggers channel opening). There is evidence for both of these mechanisms (Delmas *et al*., 2022), and as with most biological processes, the truth is likely an amalgamation of both mechanisms depending on the circumstance. For the purposes of this work, the force from lipids is the more relevant mechanism, as it can be mimicked by lateral membrane tension in MD simulations (De Vecchis *et al*., 2021) and the structural effects of this tension have been characterised experimentally (Yang *et al*., 2022).

The putative mechanism of force sensing is that mechanical stimuli to the membrane results in local deformation and flattening of the membrane dome around PIEZO1 (Guo and MacKinnon, 2017; Lin *et al*., 2019; De Vecchis *et al*., 2021). This membrane flattening releases the potential energy held in the membrane dome, which is propagated to the central region of

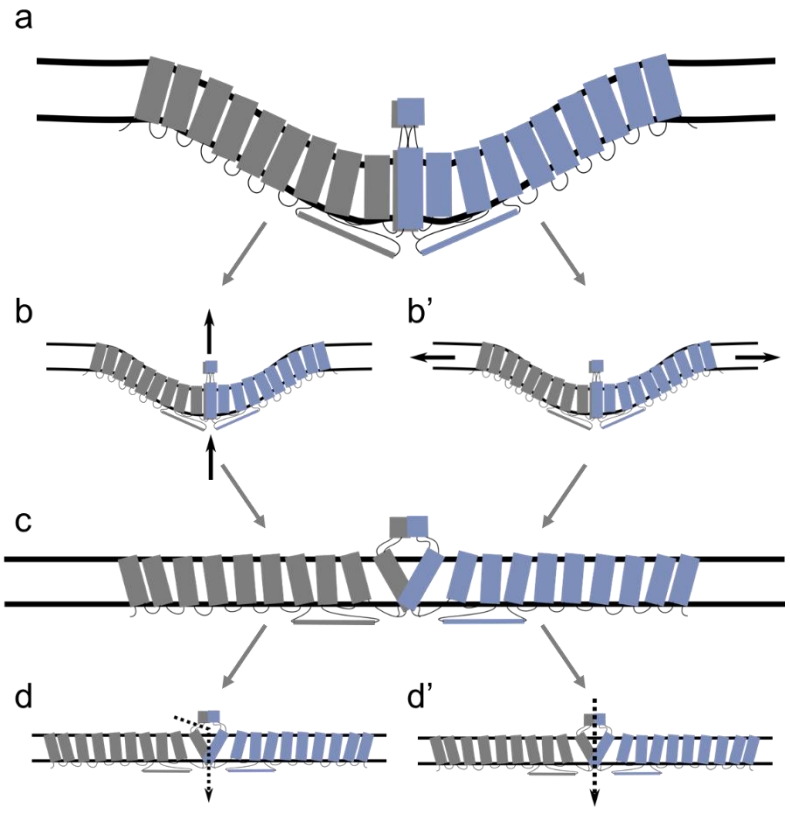


Figure 2.3 The Mechanism of Force Sensing and Activation of mPIEZO1
The curved blades of the protein form a dome in the membrane that projects into the cytosol (a) and upon membrane tension, either through interactions with extra- or intracellular tethers (i.e. force from filaments) (b) or deformation of the membrane and lateral tension (i.e. force from lipids) (b') the protein blades and dome flatten in the Z plane (c) and ion permeation through lateral fenestrations (d) or *via* the extracellular cap (d') takes place.

PIEZO1 to allow opening and ion conductance. It is likely that this force is transduced to the pore *via* the elbow region and rearrangement of the extracellular cap domain (Ge *et al*., 2015; Saotome *et al*., 2018; Zhao *et al*., 2019; Lewis and Grandl, 2020), both of which are linked to conformational change of the blades. Ions are then able to enter the pore through either the extracellular cap domain (Wang *et al*., 2019; Lewis and Grandl, 2020) or lateral fenestrations beneath it (Geng *et al*., 2020), which may occur in tandem or following the opening of hydrophobic gates in the pore and perhaps the rearrangement of a plug-and-latch domain below the pore to allow ion conductance (Geng *et al*., 2020). Following ion permeation, the channel undergoes inactivation, which appears to be controlled by subdomains in the extracellular cap (Lewis and Grandl, 2020) and residues within the pore (Bae *et al*., 2013b). Disruption of these leads to slow inactivation and thus prolonged cation influx to the cell, which causes dehydrated hereditary stomatocytosis (Bae *et al*., 2013a).

2.3.3 Lipid Interactions

Interaction with lipids is a key aspect of the physiological function of PIEZO1, as formation and deformation of the membrane dome is essential to mechanosensation, but recent experimental and structural evidence implicates specific interactions with lipids in the function and modulation of the ion channel.

2.3.3.1 Formation of a Membrane Dome

The membrane dome formed by PIEZO1 that protrudes into the cytoplasm of the cell has already been discussed here in terms of its importance for function, but the PIEZO-specific architecture and regulation of the dome are of specific interest. The first structural studies of PIEZO1 identified and modelled the membrane dome as a ~6 nm deep bowl-like structure (Guo and MacKinnon, 2017), but subsequent simulations have identified that the dome has a trilobed topology with areas of high and low curvature (Chong *et al*., 2021). Within the compartment formed by the curved blade and the back of the next the dome is the deepest, but the back of the distal HBUs produce levels of reduced curvature. Alongside this complex topology, the dome does not end at the distal region of the protein, as seen in liposome studies (Guo and MacKinnon, 2017; Yang *et al*., 2022), but extends beyond the tips of the blade and into the surrounding membrane (Chong *et al*., 2021). The differences between the structural and simulation studies are likely due to the size of the liposomes (~10 nm radius) not providing enough surface area for complex dome architecture, and potentially deforming the flexible N-terminal of the blade which likely contributes to the extension of the dome.

The formation of the dome around PIEZO in simulation allows unique linking of experimental outcomes to structural mechanisms. For instance, Ridone and colleagues (2020) demonstrated that addition of the cholesterol-removing agent methyl-β-cyclodextrin slowed

activation and inactivation of the PIEZO1 channel through its direct action on the membrane properties, rather than through protein mediated effects on cholesterol and membrane stiffness (Qi *et al*., 2015). A year later, Chong and colleagues (2021) computationally linked membrane cholesterol concentration to dome depth, which is thought to impact PIEZO1 mechanosensitivity (Haselwandter and Mackinnon, 2018). However, addition of exogenous cholesterol to HEK293 cells expressing human PIEZO1 were inhibited by addition of the lipid (Chong *et al*., 2021), which somewhat contradicts the findings with methyl-β-cyclodextrin in both studies (Ridone *et al*., 2020; Chong *et al*., 2021). The inability to unify the effect of reduced or increased cholesterol concentration could be explained by external addition compared to removal, the need for an optimal cholesterol concentration for function, or effects on PIEZO1 clustering.

The clustering of PIEZO1 was confirmed experimentally with super-resolution microscopy (Ridone *et al*., 2020) following indications of the clustering in previous studies (Cox *et al*., 2016; Maneshi *et al*., 2018), and may be a mechanism for synchronised gating and accelerated propagation of mechanical stimuli between channels (Ridone *et al*., 2020). This has been somewhat studied in simulation (Jiang *et al*., 2021a), but the simulation protocols in which the periodic boundary conditions are exploited to allow interaction of the dome with its mirror image render this work difficult to reconcile with real-world dome dynamics. However, this clustering and far reaching perturbation of the dome affect PIEZO1 function in other ways, it might impede access to highly curved regions of bilayer, such as the filopodia (Efremov *et al*., 2022).

2.3.3.2 Specific Lipid Interactions

It is difficult to delineate the role of general lipid and bilayer properties, such as stiffness, speed of diffusion, curvature and charge, and the role of specific lipid interactions with sites on the protein, as its likely that each influences the other. However, structural studies in liposomes have confirmed that lipids to bind to specific sites on the PIEZO1 channel at both the pore (Saotome *et al*., 2018; Yang *et al*., 2022) and along the blade (Yang *et al*., 2022), both in bilayers and following delipidation. The presence of lipid binding at the pore has previously been a point of dispute in the literature, as its presence in both CG and atomistic simulations (Chong *et al*., 2021; De Vecchis *et al*., 2021; Jiang *et al*., 2021a) was attributed to a simulation artefact (Jiang *et al*,. 2021b; Jiang *et al*., 2022), but the recent lipidated pore in curved and flattened structures provided further evidence for the presence of these lipids (Yang *et al*., 2022). The pore lipid may represent another gating mechanism for PIEZO1 channels, as the insertion and removal of lipid tails from hydrophobic pockets to allow ion transduction has been seen for other mechanosensing proteins (Pliotas *et al*., 2015; Kapsalis *et al*., 2020), and

they appear to move upon flattening of PIEZO1 under lateral membrane tension in simulation (De Vecchis *et al*., 2021).

Simulations have allowed identification of the potential identity of these pore lipids, likely they are POPC, POPE or another neutrally charged component of the membrane (Chong *et al*., 2021; De Vecchis *et al*., 2021). However, simulations have also identified extensive selective interactions with cholesterol at CRAC and CARC sites (Fantini and Barrantes, 2013; Chong *et al*., 2021) and PIP_2, which might be linked to the positive inside rule for membrane proteins (von Heijne, 1992; Elazar *et al*., 2016). The specific interactions with cholesterol may play a role in the modulation of functional properties of PIEZO1, as described above, and the PIP_2 interaction has been linked to the finding that depletion of PIP_2 by TRPV1 activation reduces PIEZO1 activity (Borbiro *et al*., 2015). Additionally, although they have not been included in simulation, the role of fatty acids in modulating PIEZO1 function (Romero *et al*., 2019; Ridone *et al*., 2020), may be uncovered through studying specific protein-lipid interactions. However, due to the inability to directly link the specific interactions witnessed in simulation and structures to functional and mechanistic outcomes, the study of these interactions is impeded.

3. Aims and Objectives

3.1 General Aim

My aim was to examine the structure and function of disparate membrane proteins using computational techniques with a particular focus on understanding how general and specific protein-lipid interactions affect and in turn are affected by these characteristics. To achieve this, members of the membrane-integral pyrophosphatase family and mPIEZO1 were subjected to multi-scale molecular dynamics simulation protocols to investigate key aspects of their structure, dynamics and activity.

By studying unrelated but functionally linked proteins, I can investigate a broader range of mechanistic and structural features while streamlining the computational techniques needed to do so, as I am studying the membrane interactions and dynamics of both. The selection of these proteins may allow insight into membrane protein function on a general level.

3.2 Membrane Integral Pyrophosphatases

mPPases represent clinically relevant membrane proteins, study of which could result in novel therapeutics against malaria and other neglected tropical diseases. Several areas of the mPPase mechanism remain unclear, such as the coupling of ion pumping to pyrophosphate hydrolysis, the role of intersubunit communication, and whether this is modulated by the lipid or detergent environment, and how the activity of mPPases can be targeted by inhibitors. Therefore, the following specific objectives were chosen:

(a) Identify and characterise protein-lipid interactions.
(b) Examine the dynamics of the different stages of the catalytic cycle.
(c) Use *in silico* chemistry techniques to identify new inhibitors.

3.3 Mechanosensitive Ion Channel mPIEZO1

The PIEZO1 ion channel has been implicated in a number of human cardiovascular diseases, but due to its relatively recent discovery there are several underexplored areas of its activity. The understanding of mPIEZO1 has been improved in recent years through the determination of a series of cryo-EM structures of the channel, but the large missing portions of the channel from these structures and lack of atomic resolution has confounded the understanding of the mechanism of force sensing and propagation of this stimulus to the pore at a molecular level. Therefore, the following specific objectives were investigated:

(a) Determine the effect of blade length on mPIEZO1-membrane interactions.
(b) Determine the effect of blade length on mPIEZO1 activation through membrane tension in atomistic molecular dynamics simulations.

4. Methods and Theory

4.1 Molecular Dynamics Simulations

4.1.1 Background

Molecular Dynamics (MD) simulations are a computational technique where atoms in a defined starting configuration are each allocated a randomised velocity and, using integrated classical Newtonian mechanics, successive configurations of the system are calculated (Marrink *et al*., 2019). MD simulations are able to explore a large range of resolutions, from a quantum level description of biomolecules, to ultra coarse-grained (one bead per molecule) models, where detail is lost but the accessible time-scale and size of the system can be increased (Lindahl, 2008). For the purposes of this work, I focused on atomistic and coarse-grained (CG) representations, as integration of these techniques in a multi-scale approach can enhance both of these techniques to allow both the highly accurate atomistic detail of the system including hydrogen-bonding and conformational changes (Stansfeld and Sansom, 2011), and the acceleration of simulation time through CG representation (Marrink and Tieleman, 2013). In classical atomistic simulations, every atom of the protein, solvent and lipids is modelled as an explicit particle, the properties of which are specified by their identity using a forcefield (Best, 2019). In contrast, CG systems typically model 3-6 heavy atoms and their associated hydrogen atoms as a single particle (Marrink *et al*., 2019). These particles are then allocated different properties based on the original atom identities.

4.1.2 Force fields

A forcefield is a description of the parameters for each of the particles, bonded and non-bonded interactions, angle geometry, charges and potential energy of a system, typically determined from physicochemical experiments and quantum mechanics calculations (Best, 2019). Unlike the coordinates of the particles in the system, their parameters determined by the forcefield do not change throughout the simulation, and are able to guide their motion and interactions throughout simulation time (Guvench and MacKerell, 2008). The atomistic forcefields available for simulations of proteins in lipid and solvent are some of the most developed for MD simulation and have undergone much optimisation since their first uses (Marrink *et al*., 2019), but there are increasingly well parameterised CG forcefields available.

The majority of forcefields attempt to represent two molecular properties: bonded and non-bonded interactions. Bonded interactions include the stretching, bending and rotation of bonds, whereas the nonbonded interactions include electrostatics, Pauli exclusion principle (that no two electrons can have identical values for their quantum numbers in the same atom) and dispersion. These energy terms (E_{bonded} and $E_{nonbonded}$) can be represented as:

$$(1)\ E_{bonded} = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} k_\chi[1 + \cos(n\chi - \sigma)]$$

Where b represents the distance between atoms, k_b represents the stiffness and b_0 is the equilibrium length of the bond, while θ is the bond angle, k_θ is the stiffness and θ_0 is equilibrium geometry of the angle, χ is the value of the dihedral, k_χ is the energetic parameter determining barrier height, n is the periodicity and σ is the phase (Guvench and MacKerell, 2008). However, in some forcefields, bond constraint algorithms such as LINCS remove the need to examine bond vibration (Lindahl, 2008). And:

$$(2)\ E_{nonbonded} = \sum_{\substack{nonbonded\ pairs \\ ij}} \left(\varepsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 * \left(\frac{R_{min,ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{r_{ij}} \right)$$

The first term in this equation within square brackets, is the Leonard-Jones expression, which models dispersion and the Pauli exclusion principle. As r_{ij} represents the separation of the atoms, the first section of the term dictates that atoms are repulsed at small distances beyond an energy minimum, and the first section dictates that atoms feel increasing attraction as they move closer before reaching the energy minimum. ε_{ij} and $R_{min,ij}$ are based on the two interacting atoms i and j, representing the interaction minimum and the distance where the Leonard-Jones energy is a minimum, respectively. The latter term in the equation is Coulomb's law, in which the charge of the atoms i and j are represented by q_i and q_j, respectively, and so models the electrostatic interactions.

These equations are then combined so that the total energy (E_{total}) is:

$$(3)\ E_{total} = E_{bonded} + E_{nonbonded} + E_{other}$$

The term E_{other} represents other elements that are specific to the forcefield (Guvench and MacKerell, 2008).

4.1.2.1 CHARMM-36m

Chemistry at Harvard Molecular Mechanics or "CHARMM" is one of the main atomistic force field families and is distinct from other all-atom forcefields, due to its use of quantum mechanic energies when parametrising interactions between biomolecules and water and sidechain dihedrals (Brooks *et al*., 2009). The water model in this forcefield is a three-site water model, TIP3P, which is slightly different to other forcefields as, in addition to other parameters, the behaviour of the water model is described by a Leonard-Jones term, but has not been found to be significantly different in practice (Best, 2019). There are a number of difficulties when developing forcefields for membrane proteins (Best, 2019), due to the high viscosity of the membrane, and lack of experimental data for membrane proteins as a result of barriers to their study. Despite this, the CHARMM-36 (Huang *et al*., 2016) is thought to be well parametrised for the type of work undertaken in this thesis (Best, 2019). The CHARMM-36m forcefield can

be applied to systems through the CHARMM-GUI (Lee *et al*., 2016b), which also allows parametrisation of new molecules, although the experimental validity of these models cannot always be guaranteed.

4.1.2.2 Martini

The Martini forcefield was first developed to study the properties of lipid bilayers, as it could be used to access timescales outside the range of atomistic simulations, while maintaining the properties of the lipids in question (Bruininks *et al*., 2019). Following this initial application, the forcefield has been expanded to include proteins and other biomolecules (Monticelli *et al*., 2008). The beads available in the Martini forcefield have four chemical classes: charged, polar, nonpolar and apolar. Within in these are subtypes that determine their hydrogen-bonding capability (i.e. donor and/or acceptor or neither), or the strength of their polar properties. This mixture of classes and subtypes provides 18 different beads (4 charged, 4 nonpolar, 5 polar and 5 apolar), and there are 10 non-bonded interaction levels based on experimental data for each of the pairs of these beads (Bruininks *et al*., 2019). In contrast, the bonded interactions between these beads was defined through mapping of the beads onto atomistic simulations and simple bonded potentials (Bruininks *et al*., 2019). Due to this method of parameterisation, parametrising new molecules is a possible but time-consuming process.

4.1.3 Integration and time step

MD simulations involve the iterative solving of the equations of motion in steps known as “timesteps”, integration of these equations has to be performed at regular intervals using an integration algorithm to calculate the position and velocity of the particles in the system. Integration algorithms have to be used to approximate these terms numerically, as they cannot be solved analytically. To begin the simulation, random velocities are generated for each of the particles in the simulation using a Boltzmann distribution, which are then recalculated between each timestep to determine the coordinates of the system (Bauer *et al*., 2022). The timestep is determined by ensuring the size of step between frames of the simulation are not so small that it requires a large amount of computer resources, but not so large that particles interact inappropriately, either though not interacting at all and moving “through” each other, or by clashing. In GROMACS (Van Der Spoel *et al*., 2005), the software used in this thesis, a Leap-Frog integration method is used, where the calculation of the velocities and the coordinates “leap” over each other half-step intervals (Bauer *et al*., 2022).

4.1.4 Periodic boundary conditions

MD simulations often use periodic boundary conditions, which describe the ability of a particle exiting the simulation box to re-enter at the opposite side, as hard boundaries are unnatural in biological systems (Bauer *et al*., 2022), although these can be exploited in some

circumstances. This is essential, as it reduces surface artifacts, edge effects and allows for infinite simulation space using a set number of atoms (Lindahl, 2008). The only accompanying requirements for periodic boundary conditions, is that the simulation box is sufficiently large, there are appropriate cut-off values for nonbonded forcefield parameters to prevent mirror images of the simulation from interacting with each other and the box shape perfectly tessellates, to ensure there are no gaps with associated vacuum effects. The size of the simulation is directly linked to computational time, and the use of cut-offs can be detrimental, leading to drifts in energy and heating up of the system (Lindahl, 2008). In the case of the software used in this work, GROMACS (Van Der Spoel *et al*., 2005), only the nearest image of each particle is used for long-range interactions.

4.1.5 Statistical Mechanics and Thermodynamic Ensembles

Statistical mechanics is a method of explaining the behaviour of a system through applying statistical and probability methods to an ensemble of particles, and thereby linking macroscale properties such as temperature and pressure to the parameters of microscopic particles. In this work, the statistical mechanics are based in classical thermodynamics, rather than quantum statistics. In statistical mechanics, a thermodynamic ensemble represents the possible states of a system, often representing equilibrium. There are three major ensembles used in Molecular Dynamics simulations:

1. NPT (isothermal-isotonic ensemble): the number of particles, pressure and temperature are maintained.
2. NVT (canonical ensemble): the number of particles, volume and temperature are maintained.
3. NVE (microcanonical ensemble): the number of particles, volume and energy are maintained.

Each of these can be used to simulate a system, but the NPT ensemble was used for the production simulations presented in this work, as the energy of the system must change as a result of conformational changes to the system, and the pressure needs to be maintained to ensure particle interactions or allow creation of lateral tension in the atomistic mPIEZO1 simulations.

4.1.6 Thermostat and barostat

Thermostats and barostats are standard methods of prevent the temperature and pressure of the system from changing during simulation time. These scale the velocities during integration of steps to maintain the set temperature, or scale the box size in the XY and Z directions to maintain the set pressure (Lindahl, 2008). By maintaining the temperature and pressure, the

total energy of the system is not conserved. The thermostat and barostats for each simulation are defined below.

4.1.7 Protein Structure Prediction and Modelling

In recent years, there has been an advent of protein structure prediction and modelling, to bypass barriers to protein structure determination through experimental methods and exploit the advances in computational power (Lushington, 2014). There are three major methods of protein structure prediction: comparative or homology modelling, where an appropriate template structure is selected and the target is modelled based on structural alignments, *ab initio* or *de novo* modelling, in which the model is constructed based on searching the conformational space of the protein sequence, and deep learning-based methods, where a neural network is trained on the existing structures in the Protein Data Bank to model a structure based on a provided amino acid sequence. Critical Assessment of Structure Prediction (CASP) has been taking place since 1994, in which every 2 years, prediction methods are tested against recently determined protein structures to assess their performance. Prior to 2018, the highest ranked methods were primarily comparative modelling and *ab initio* by Robetta (Song *et al.,* 2013), but now deep learning-based methods dominate the results, such as AlphaFold2 (Jumper *et al.,* 2021). In this work, comparative modelling through Robetta and AlphaFold2 have been used to generate potential structural models, as will be discussed later in this chapter.

4.2 Protein Preparation

The choice of protein structure for simulation has many considerations encompassing quality (resolution, completeness, fitting to the electron density map, modelling errors) as discussed by (Arkhipova *et al*., 2017), the biological relevance of the structure (catalytic or conformational state, mutagenesis, bound ligands and ions) and the hypothesis being addressed through simulation. Despite potential limitations of experimentally determined protein structures, a number of these issues can be resolved through appropriate preparation of the structure. For example, missing regions can be modelled, amino acid sequences can be converted to their wild-type identity and side chains can adopt different orientations. Additionally, energy minimisation and equilibration processes can eliminate atom clashes. The energy minimisation method used in this work was the steepest decent algorithm, in which the atom coordinates are refined iteratively, where each step is in the steepest direction possible at that point to a local energy minimum. These techniques can allow further exploration of the biological space without significant limitations of the structure availability and quality. Despite this, the quality of the underlying model must be assessed and determined appropriate for simulation before continuing (Arkhipova *et al*., 2017).

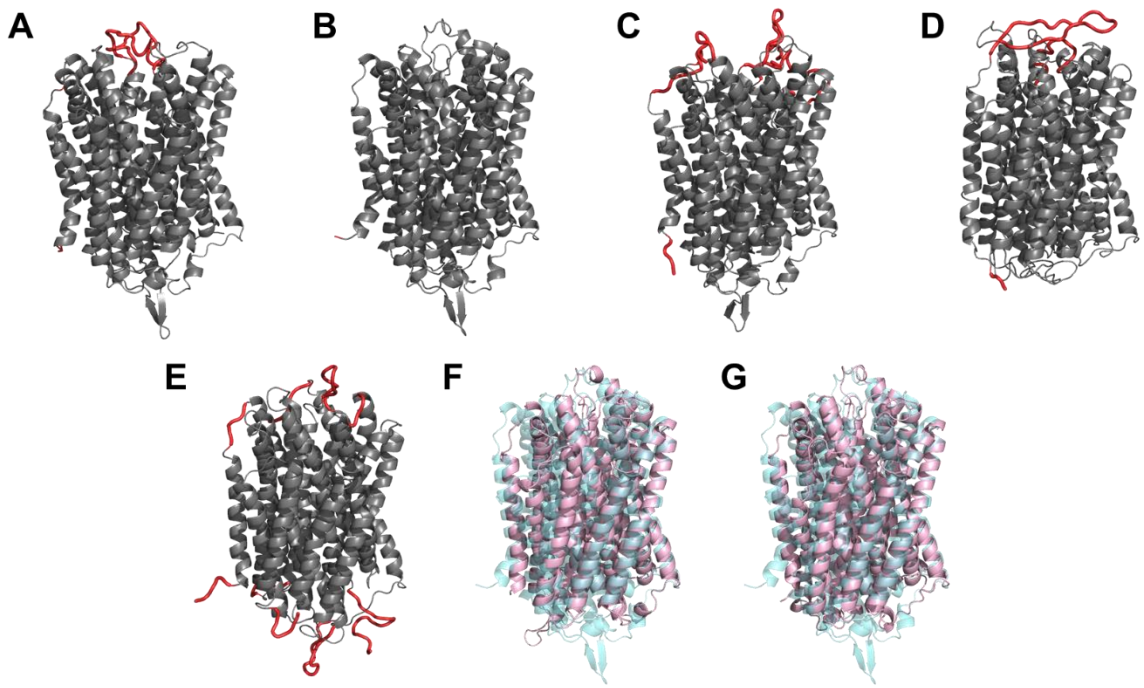


Figure 4.1 The modelling of unresolved regions of mPPase structures and *Cp*-PPase
The regions of the mPPase structures added through Modeller are shown as red tubes on the grey resolved regions of **A)** the resting state (PDB: 4AV3), **B)** substrate-bound state (PDB: 5LZQ), **C)** product-bound state (PDB: 4AV6) of *Tm*-PPase, **D)** relaxed product-bound state of *Vr*-PPase (PDB: 5GPJ) and **E)** the substrate-bound state of *Pa*-PPase (unpublished). The homology modelled structures of Cp-PPase as modelled by **F)** Robetta (pink) and **G)** AlphaFold2 (pink) compared with the *Vr*-PPase structure (cyan) (PDB: 5GPJ).

4.2.1 mPPase Protein Preparation

All of the experimentally resolved mPPase structures used in this work were obtained through x-ray crystallography (**Table 4.1**), a technique that has been used extensively for membrane protein structure determination (Kermani, 2021). However, it has come under a lot of scrutiny, as the processes involved in crystallography can detrimentally alter the protein structure, stability and obscure interactions with native ligands and lipids (Bill *et al*., 2011). Despite this, it continues to be the primary mechanism for solving protein structures (Protein DataBank, 2022) and has led to several key mechanistic insights into mPPase function.

The coordinates of the protein structures were accessed through the Protein Data Bank (rcsb.org) (Berman *et al*., 2000), or in the case of *Pa*-PPase, these were provided by collaborators in PDB format (Strauss, 2021). These were then initially prepared for simulation following standard protocols (Biggin and Bond, 2008), including removal of all non-protein atoms, bound ligands and ions in the catalytic centre of the proteins. Any unresolved portions of the protein were added using Modeller (Webb and Sali, 2016) (**Fig. 4.1**), as detailed in **Table 4.1**, and any residue substitution mutations were generated through the same means (**Table 4.2**). Additionally, to model the 5-6 loop open on the substrate bound structure, the atoms were removed from the structure and remodelled using Modeller. The atom names and format were processed to allow simulation with GROMACS through the pdb2gmx programme

mPPase	Catalytic state	Resolution (Å)	Modelled regions	PDB and reference
Tm-PPase	Resting state	2.6	1, 30, 211-221, and 577-595	4AV3 (Kellosalo *et al*., 2012)
	Substrate-bound state	3.5	1	5LZQ (Li *et al*., 2016)
	Product-bound state	4.0	1-4, 29-37, 114-119, 208-225, 473-479, and 572-599	4AV6 (Kellosalo *et al*., 2012)
Vr-PPase	Relaxed product-bound state	3.5	1-3, 39-62, and 262-278	5GPJ (Li *et al*., 2016)
Pa-PPase	Substrate-bound state	3.8	1-5, 31-35, 74-78, 159-164, 201-211, 317-322, 508-520, 575-580, 658-661 and 706-721	-

Table 4.1 Summary of the crystallographic mPPase structures used in this work

using the OPLS-AA all atom forcefield (Jorgensen *et al*., 1996). The resulting protein models were visually inspected for errors or knotted loops prior to specific processing dependent on the mode of simulation (i.e. CG or atomistic).

4.2.2 Homology Modelling of *Cp*-PPase

The work outlined in section 5 required simulation of an mPPase structure that had not yet been experimentally determined. The protein chosen was *Cp*-PPase, as it is a crystallographic target and has been previously used for homology modelling, which resulted in an mPPase-like model (Harborne *et al*., 2020). However, in the same ways that experimentally derived models should be assessed for their quality, the quality of predicted structures should also be checked (Benkert *et al*., 2011). Upon inspection of the electrostatic profiles of these homology models generated by iTasser (Yang *et al*., 2014) and Modeller (Webb and Sali, 2016) indicated that they may not be as comparable to other mPPase structures as previously thought. Therefore, I decided to generate new homology models for this work that may allow better predictions of mPPase behaviour. The *Cp*-PPase amino acid sequence was submitted to the Robetta (robetta.bakerlab.org) server (Song *et al*., 2013) and AlphaFold2 (Jumper *et al*., 2021)

for homology modelling. Both of these identified the *Vr*-PPase relaxed product-bound state (PDB: 5GPJ) as the best template for homology modelling.

Protein	Name	% Anionic Lipid	Mutations	CG simulations (μs)
Tm-PPase	TmPPE100	0	-	5x 5
	TmMix10	10 each[a]	-	5x 5
	TmPA20 TmPG20 TmPS20	20	-	5x 5
	TmPA20_SSM TmPG20_SSM TmPS20_SSM		$K^{9.70}A$ $K^{10.49}A$ $K^{13.52}A$ $K^{14.48}A$[b]	5x 5
	TmPA20_DSM TmPG20_DSM TMPS20_DSM		$K^{9.70}A$ $K^{10.49}A$ $K^{13.52}A$ $K^{14.48}A$	5x 5
	TmPA40 TmPG40 TmPS40	40	-	5x 5
Vr-PPase	VrTonoplast	12 cumulative	-	4x 5 1x 20
	VrTonoplast_DSM		$K^{10.49}A$ $K^{13.48}A$ $K^{14.40}A$ $K^{14.48}A$	5x 5
Cp-PPase	CpMix10	10 each[a]	-	5x 5
	CpPA20 CpPG20 CpPS20	20	-	5x 5
	CpPA20_DSM CpPG20_DSM CpPS20_DSM		$K^{9.69}A$ $K^{9.73}A$ $R^{13.52}A$ $R^{14.48}A$	5x 5

Table 4.2 The mPPase CG simulations
[a]here the membrane was composed of equal numbers POPA, POPG and POPS
[b]these mutations were to a single interfacial interaction site

The electrostatic potential of both the Robetta and AlphaFold2 models matched what was anticipated and had been seen previously for other mPPases, and so they were submitted for model quality analysis through SWISS-MODEL Tools (swissmodel.expasy.org/qmean) for QMEAN (qualitive model energy analysis) and Z-score analysis using the QMEANBrane

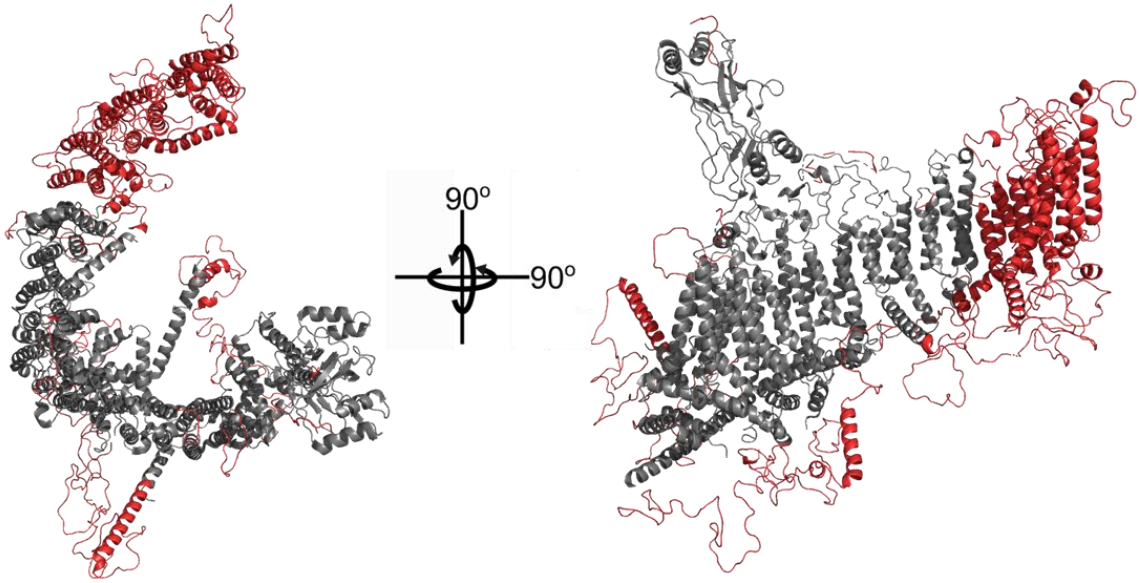


Figure 4.2 The modelling of unresolved regions of mPIEZO1
The regions of the mPIEZO1 structure added through Modeller are shown in red against the grey resolved regions of the mPIEZO1 structure (PDB: 6B3R).

option (Studer *et al*., 2014). This confirmed that the homology models generated in this work are higher quality than those previously generated through iTasser and Modeller. Despite similar structures being generated, the Robetta model was chosen for simulation in this work as the arrangement of the helices in this model more closely resembled the *Vr*-PPase template.

4.2.3 mPIEZO1 Protein Preparation

So far, all of the PIEZO1 and PIEZO2 protein structures have been resolved using cryo-electron microscopy, which has been increasingly used in recent years to resolve the structure of complex membrane protein structures (García-Nafría and Tate, 2020; Kühlbrandt, 2022). This technique overcomes some of the previously mentioned x-ray crystallography limitations, by being more compatible with detergent alternatives, not relying on long-term stability for crystallisation and allowing multiple conformations to be identified and sampled at once (Kampjut *et al*., 2021). Despite this, there are still some limitations to the use of cryo-EM for membrane protein structure determination, such as reduced resolution, technical issues that can prevent full structure sampling such as preferred orientations and protein size limits (Renaud *et al*., 2018). However, for PIEZO structural study, the technique has been revolutionary, as it has also allowed study of PIEZO-induced curvature of membranes (Guo and MacKinnon, 2017) and pressure induced flattening (Yang *et al*., 2022).

Despite how useful cryo-EM has been, a major issue with the available mPIEZO1 structures has been the lack of cytoplasmic loops and the N-terminal helical bundles (**Fig. 4.2**). Previous work in the Kalli research group has addressed this, through building a model of the full length mPIEZO1 and demonstrating how the full-length structure alters membrane dome formation

and lipid interactions (Chong *et al*., 2021). The process by which this full-length structure was generated has been described previously (Chong *et al*., 2021). In short, the 4-5-6 bundles of the truncated mPIEZO1 structure (PDB: 6B3R) (Guo and MacKinnon, 2017) were used as templates for structure prediction using MEMSAT-SVM (Nugent and Jones, 2012), PSIPRED (Jones, 1999; Buchan and Jones, 2019) and Modeller (Webb and Sali, 2016). In this work, I have used this full-length mPIEZO1 to create several truncated versions of the protein by removing the relevant atoms in PyMol (DeLano, 2002), exporting the resulting structure and processing through the pdb2gmx programme, as before.

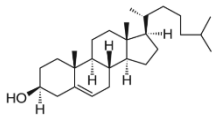
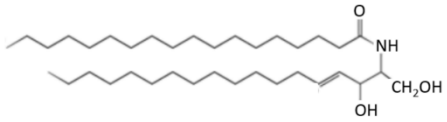
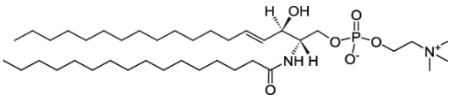
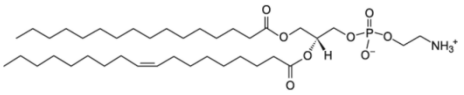
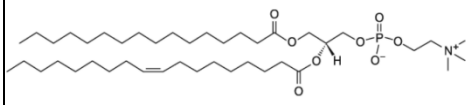
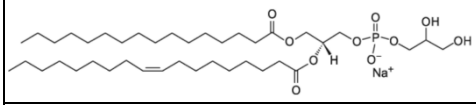
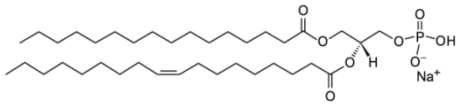
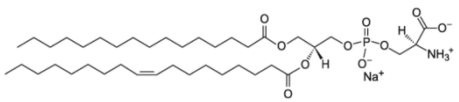
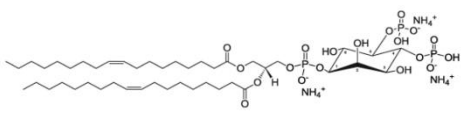
Lipid	Type	Diagram	Tails	Charge
Cholesterol	Sterol		C(d18:1/ 18:0)	0
Ceramide hexoside	Sphingolipid		C(d18:1/ 18:0)	0
Sphingomyelin			C(d18:1/ 18:0)	0
Phosphatidyl-ethanolamine	1-Palmitoyl-2-Oleoyl		C16:0/ 18:1	0
Phosphatidyl-choline				0
Phosphatidyl-glycerol				-1
Phosphatidic acid				-2
Phosphatidyl-serine				-1
Phosphatidy-linositol bisphosphate	Phosphatidy-linositol		C16:1(9c) C18:1(9c)	-5

Table 4.3 Details of the lipids used in these simulations

4.3 Coarse-Grained Simulations

All CG simulations were performed using the MARTINI 2.2 forcefield (Monticelli *et al*., 2008) and GROMACS 5.0.X (www.gromacs.org) (Van Der Spoel *et al*., 2005). The protein structures were converted to MARTINI CG models using martinize.py (De Jong *et al*., 2013) and an elastic network was applied to all protein backbone atoms within 0.7 nm with a force constraint of 1000 $KJmol^{-1}nm^2$ to maintain the secondary and tertiary structure of the protein, as well as merging the subunits to maintain their interfaces. The resulting structures were energy minimised in a vacuum using the steepest descent algorithm in GROMACS. The mPPase CG proteins were centred in 16 x 16.5 x 16 nm^3 simulation boxes, whereas the mPIEZO1 CG models were centred in 44 x 44 x 24 nm^3 simulation boxes.

Name	Mutations	Bilayer size (nm^2)	CG Simulations (μs)
$mPIEZO1_{full\text{-}length}$	-	44 x 44	3x 3
		54 x 54	1x 1
$mPIEZO1_{Beam}$	1300 – 1365 replaced by reversed sequence of 1366 – 1431	44 x 44	3x 3
		54 x 54	1x 1
$mPIEZO1_{\Delta HBU1}$	1 – 175 removed	44 x 44	3x 3
$mPIEZO1_{\Delta HBU1\text{-}2}$	1 – 400 removed	44 x 44	3x 3
$mPIEZO1_{\Delta HBU1\text{-}3}$	1 – 550 removed	44 x 44	3x 3
$mPIEZO1_{\Delta HBU1\text{-}4}$	1 – 770 removed	44 x 44	3x 3
$mPIEZO1_{\Delta HBU1\text{-}5}$	1 – 960 removed	44 x 44	3x 3
$mPIEZO1_{\Delta HBU1\text{-}6}$	1 – 1140 removed	44 x 44	3x 3
$mPIEZO1_{\Delta HBU1\text{-}7}$	1 – 1365 removed	44 x 44	3x 3
$mPIEZO1_{\Delta HBU1\text{-}8}$	1 – 1970 removed	44 x 44	3x 3

Table 4.4 The CG simulations of mPIEZO1

Following this, lipids (**Table 4.3**) were assembled around the protein and the system solvated by random placement of water and NaCl (150 mM) using the insane method (Wassenaar *et al*., 2015). The mPPase CG systems are outlined in **Table 4.2**. In brief, the *Tm*-PPase and *Cp*-PPase systems were simulated in symmetric bilayers with one predominant neutral lipid (POPE) and either one or three anionic lipids (POPA, POPG or POPS), whereas the *Vr*-PPase simulations were performed in a bilayer resembling the *Vigna radiata* tonoplast membrane (cholesterol (29%), POPC (25%), POPE (17%), ceramide hexoside (17%), PIP_2 (6%), POPG (3%), POPS (2%), POPA (1%)) (Yoshida and Uemura, 1986). The mPIEZO1 simulations (**Table 4.4**) were performed in asymmetric bilayers resembling the endothelial cell membrane

(upper leaflet: POPC (55%), DPSM (5%), POPE (20%), CHOL (20%), lower leaflet: POPC (50%), POPE (20%), POPS (5%), CHOL (20%), POP2 (5%)).

A further energy minimisation step was performed after lipidation and solvation of the systems and an equilibration with position restrains on the backbone protein particles took place. For the mPPase systems, this was performed at 323 K for 2 ns, and the final frame was used to generate 5 independent repeat production simulations, each for 5 μs, where a V-rescale thermostat (323 or 310 K) and Parrinello-Rahman barostat (1 bar) was used to maintain the temperature and pressure of the system. The mPIEZO1 systems underwent 500 ns of equilibration at 310 K to allow adequate time for the membrane dome to form around the protein. During formation of the membrane dome, some lipids would be trapped by the protein or flipped from one membrane leaflet to the other, and so had to be identified and removed. The final frames were used to generate 3 independent production simulations performed at 310 K for 3 μs each. The same production simulation parameters were used as for the mPPase simulations, but in some cases, a lower integration timestep of 2 fs was used.

4.4 All Atom Simulations

4.4.1 Backmapping

Throughout this work, a serial multi-scale approach was used, in which representative final frames of selected CG simulations were converted to atomistic representations, colloquially known as “backmapping”. This conversion used the Backward protocol, as described previously (Wassenaar *et al*., 2014), where the lipids and amino acids of the final CG snapshot are reconstructed and the resulting converted systems are energy minimised and equilibrated for 2 ns. For successful mPIEZO1 backmapping, some of the generated atoms had to be manually moved by a maximum of 0.1 Å to allow energy minimisation to proceed. For each mPPase backmapped simulation (**Table 4.5**), 3 replicas were generated with random starting velocities that were simulated using the CHARMM-36m forcefield (Huang *et al*., 2016), a Nosé-Hoover thermostat and a 2 fs time step.

The mPIEZO1 backmapped systems required further processing prior to production simulation. Firstly, the water and ions were removed from the systems and the simulation box size increased in the Z axis by a minimum of 40%. These new boxes were solvated and any water molecules overlapping with the membrane and protein were removed prior to energy minimisation. Calcium ions were added to the system by random replacement of 50 water molecules. This was also used to add enough sodium and chloride ions to neutralise the system and provide a 150 mM NaCl concentration. This neutralised system was further energy minimised and equilibrated using three independent NPT equilibration steps totalling 1.035 ns, followed by a 5 ns equilibration step. These systems were simulated under pressure

between 1 and -40 bar on the membrane in the XY plane, which was maintained by a Parrinello-Rahman barostat and at 310 K for a minimum of 50 ns (**Table 4.6**).

Name	Mutations (both subunits)	Backmapped time (ns)
TmPA20 TmPG20 TmPS20	-	3x 250
TmPA20_DSM TmPG20_DSM TMPS20_DSM	$K^{9.70}A$ $K^{10.49}A$ $K^{13.52}A$ $K^{14.48}A$	3x 250
VrTonoplast	-	3x 250
VrTonoplast_DSM	$K^{10.49}A$ $K^{13.48}A$ $K^{14.40}A$ $K^{14.48}A$	3x 250
CpPA20 CpPG20 CpPS20	-	3x 250
CpPA20_DSM CpPG20_DSM CpPS20_DSM	$K^{9.69}A$ $K^{9.73}A$ $R^{13.52}A$ $R^{14.48}A$	3x 250

Table 4.5 The backmapped mPPase simulations

Name	Box size (nm)	Force (bar)	Backmapped time (ns)
$mPIEZO1_{WT}$	44 x 44 x 34	1	1x 50
		-20	1x 50
	44 x 44 x 42	-35	1x 50
	54 x 54 x 42	-40	1x 50

Table 4.6 The backmapped mPIEZO1 simulations

4.4.2 Atomistic

In addition to a serial multi-scale approach, the atomistic catalytic simulations of *Tm*-PPase outlined in chapter 4 were performed solely as all atom simulations to avoid any loss of subtle structural features during the CG simulation and conversion processes, and due to the lack of any CG simulation parameters for the small molecules in these simulations. Following protein preparation, the structures were embedded in a POPE (80%) and POPG (20%) bilayer, solvated with TIP3 water model and NaCl (150 mM) using the CHARMM-GUI Input Generator (Lee *et al*., 2016b). The simulation boxes were ~14.5 nm^2 in the membrane plane and 12 nm^2 in z. The small molecules and ions in complex with *Tm*-PPase were added back to their relevant structures in the same position as found in the crystal structures. However, some substitutions were made, such as replacement of the Ca^{2+} in the resting state structure with a

State	Symmetry	Ligands/Ions	Time (ns)	Other
Resting state	Symmetrical	2 Mg^{2+}	2x 250 1x 750	Active site Ca^{2+} substituted for Mg^{2+}
		5 Mg^{2+} 1 IDP 1 Na^{+}	3x 100	-
		5 Mg^{2+} 1 IDP	3x 250	-
Substrate-bound state	Symmetrical	5 Mg^{2+} 1 IDP 1 Na^{+}	3x 500	-
		5 Mg^{2+} 1 ETI 1 Na^{+}	3x 250	-
		5 Mg^{2+} 1 ETI 1 Na^{+}	3x 250	Alternate ETI positioning
		2 Mg^{2+}	2x 500 1x 750	“empty” substrate-bound state
		2 Mg^{2+}	3x 250	Loop modelled open
		-	3x 100	-
		7795842 2 Mg^{2+}	3x 100	-
	Asymmetrical	5 Mg^{2+} 1 IDP 1 Na^{+}	3x 250	-
		5 Mg^{2+} 1 ETI 1 Na^{+}	3x 250	-
		5 Mg^{2+} 1 ETI 1 Na^{+}	3x 250	Alternate ETI positioning
Product-bound	Symmetrical	2 PO_4^{-} 1 K^{+} 4 Mg^{2+}	3x 250	-

Table 4.7 The atomistic mPPase simulations

second Mg^{2+}, to be physiologically relevant as Ca^{2+} ions are inhibitory (Kellosalo *et al*., 2012). The IDP in the substrate-bound crystal structure was also used to model another non-hydrolysable substrate analogue, ETI, in the hydrolytic centre, by aligning the small molecules.

Due to adding back these non-protein molecules, several equilibration strategies were tested to maintain substrate positioning but allow some equilibration of their positioning. The most successful protocol involved a stepwise reduction in position restraints, from 1000 $KJmol^{-1}nm^2$ for 2 ns to 500 $KJmol^{-2}nm^2$ for the small molecules and Mg^{2+} ions but maintained at 1000 $KJmol^{-1}nm^2$ for the C_α protein atoms, for a further 10 ns. The resulting configuration was used to generate 3 independent repeats simulated for up to 750 ns simulation time (**Table 4.7**).

4.5 Analysis

4.5.1 General

The trajectories were prepared post simulation for analysis by concatenating all parts and both centring and fitting the protein in the centre of the simulation box. To normalise the number of frames in the mPIEZO1 simulations where the integration timestep was reduced, the outputted frames were skipped to match with the unmodified simulations, any other timestep skipping is detailed in the specific analysis detail. All analyses were performed using GROMACS, VMD (Humphrey *et al*., 1996), PyMol (DeLano, 2002), PyLipID (Song *et al*., 2022b) and locally written scripts.

4.5.2 Lipid contact analyses

To quantify and identify lipid contacts with the protein, a locally written script was used to calculate the interactions between each backbone residue of the protein and the lipid head groups or the entirety of the cholesterol molecule in the simulations. A cut-off of 5.5 Å or 4 Å was defined for CG or AT simulations, respectively. The raw number of contacts was normalised in relation to the total number of simulation frames and the number of that lipid specific in the bilayer. The GROMACS tools gmx densmap and gmx mindist were used in combination with the contacts to further understand the lipid binding. The densmap tool was used to visualise the density of the headgroups of the lipids around the fitted protein structure, or the entirety of the cholesterol molecule, and for normalisation the replicate trajectories were concatenated. The mindist tool was used to understand the number of contacts between lipid binding sites and the lipids in the bilayer. The binding site was treated as a group and the cut off matched those used to define lipid contacts. To further characterise the lipid binding sites identified in the mPPase work, PyLipID was used with the cut-offs 0.475-0.8 with a skipped trajectory to understand lipid residence time at each mPPase residue. In addition to these tools, the membrane curvature induced by mPIEZO1 was analysed using a locally written depth script (Chong *et al*., 2021).

4.5.3 Protein dynamics and structure

A number of GROMACS tools were used to understand the dynamics of the simulated proteins, such as the RMSD to study the structural changes over time, the RMSF to understand the dynamics of different protein regions and densmap to visualise protein motion. In addition to these, gmx bundle and gmx gangle tools have been used to study the motions of the mPIEZO1 helical bundles in relation to each other and the Z axis. Additionally, locally written scripts assessing protein contacts were used in combination with gmx mindist to understand how protein subunits interacted with one another. These were processed in the same ways as the protein-lipid contacts scripts, but with 1.5 nm contact cut-offs.

Aspects of the protein structures were assessed using static structures, such as the electrostatic profiles, which were calculated by preparing the structure using PDB2PQR (Dolinsky *et al*., 2004) and then using the APBS plugin for VMD (Baker *et al*., 2001). Additionally, the proteins were aligned using both VMD and PyMol to visualise and measure movements during simulation. PyMol was also used to inspect the electron density of the available structures and their bound lipids, ligands and ions.

4.6 Computational Chemistry Techniques

4.6.1 High-Throughput *in silico* Screening

The post-simulation structures of the *Tm*-PPase resting state and substrate-bound state with small molecules removed and loop modelled open (“empty” substrate-bound state) were prepared for *in silico* screening using Maestro 2017.3 (Halgren *et al*., 2004). The Protein Preparation Wizard was used for preprocessing at pH 7 ± 0.2. Initially SiteMap was used to identify druggable sites on the protein, but the active site was of most interest, and so the 10 x 10 x 10 $Å^3$ inner grid box (and 40 x 40 x 40 $Å^3$ outer grid box) was centred on the active site Mg^{2+}.

Both models underwent virtual screening with the internal MCCB library of 91,000 compounds using both Glide (Friesner *et al*., 2004) and Gold (Jones *et al*., 1997) programmes. These ligands were also prepared in Maestro at pH 7 ± 0.2 and the OPLS forcefield. From the Glide HTVS screen, the top 5000 compounds were ranked by docking score and redocked using standard precision. Following this, the top 100 as ranked by docking score were redocked with the same precision and inspected for their likelihood of representing a good candidate molecule for functional testing. This resulted in a shortlist of 20-30 compounds, which were docked with extra precision and up to 10 binding poses outputted for each. These were then further analysed to identify the compounds with favourable docking scores, conformational consensus across their docking poses and favourable interactions, particularly with the metal ions.

4.6.2 Compound Optimisation of 7795842

The compound 7795842 was identified through *in vitro* investigation as a good candidate for compound optimisation, as it had the highest activity and is commercially available. Compound 7795842 was considered as 2 groups: the “cephalosporin” group and the “acid” group, both of which would be used to build a SAR library. The acid group was used for ROCS (rapid overlay of chemical structures) analysis in combination with the Sigma-Aldrich library to identify acid groups with similar strictures to compound 7795842. From the top 250 identified chemical structures, the most relevant based on the structure and functional groups were selected and shortlisted according to their commercial availability, resulting in 73 acids. These were then computationally converted to amides using Maestro and docked into the “empty” substrate-bound model using extra precision Glide docking. 57 of these were then taken forward for extra precision docking with 10 outputted docking poses, resulting in similar docking scores to the original compound 7795842.

4.6.3 *In vitro* Compound Screening

The *in vitro* screening was performed by Dr Keni Vidilaseris in Helskini, as described previously (Vidilaseris *et al*., 2018). The activity assay is a colourimetric method based on the molybdenum blue protocol to measure the conversion of PP_i to phosphate ions by the mPPase. The compound of interest was solubilised in DMSO and diluted to the required concentration (100 µM for initial activity screening, or 0.01, 0.1, 1, 5, 10, 50, 100 and 500 µM for IC_{50} investigation). The *Tm*-PPase protein was purified from *Saccharomyces cerevisiae* and reactivated for the assay by addition of 30 mg/ml soybean lecithin (40 µM) in 20% DDM (22.5 µM) that was heated to 55 °C for 15 mins and then cooled to room temperature before addition to the protein (13 mg/mL). The protein was diluted to 1/75 in the reaction mixture (0.2 M Tris-Cl pH 8.0, 8 mm $MgCl_2$, 0.33 M KCl, 0.67 M NaCl) and incubated with the inhibitors for 5 minutes at 71 °C. Following this, 2 mM Na_2PP_i (for 0.4 mM final concentration) was added to the reaction for a further 5 mins of incubation. The reaction was then cooled on ice for 5 mins, centrifuged to collect the reaction at the bottom of the tube and returned to the ice. Solution A+B (A: 3% w/v ascorbic acid in ice cold HCl (0.5 M), B: 0.7% w/v ammonium heptamolybdate in ice cold water) was added to the reaction and incubated for 10 mins. Arsenic solution (2.3 mM Na.Arsenite, 5.88 mM trisodiumcitrate.$2H_2O$ in 2% acetic acid in water) was then added and the reaction and after an hour at room temperature, the absorbance of the solution was read at 860 nm. As controls, phosphate (0 nmol, 2.5 nmol, 10 nmol and 20 nmol) was prepared in reaction mixture without protein addition, as well as substituting the compound for water, IDP (100 µM), MTI61 (200 µM) and AKI XVII103 (10 µM).

5. mPPases Create a Conserved Anionic Membrane Fingerprint as Identified *via* Multi-Scale Simulations

This chapter is based on the following publication cited in this thesis as (Holmes *et al*., 2022) Holmes *et al*., mPPases Create a Conserved Anionic Membrane Fingerprint as Identified via Multi-Scale Simulations. PLOS Comp Biol. 2022. doi: 10.1101/2022.03.08.483421.

5.1 Introduction and Rationale

Membrane integral pyrophosphatases (mPPases) are a family of membrane proteins responsible for coupling the hydrolysis of the pyrophosphate (PP_i) phosphoanhydride bond to the pumping of a cation across the membrane (Holmes *et al*., 2019). This allows mPPases to both remove excess PP_i from the cytoplasm, and to generate a membrane potential. mPPases are found in all kingdoms of life, excluding fungi and multicellular animals (Kajander *et al*., 2013). Due to this, they are validated selectively toxic drug targets against a variety of protozoan and bacterial pathogens, and so inhibiting the function of mPPases may be a method of combatting these infectious pathogens (Rodrigues *et al*., 2000; Lemercier *et al*., 2002; Yoon *et al*., 2013; Liu *et al*., 2014).

Crystal structures revealed that mPPases exist as homodimers, where each subunit is composed of 16 transmembrane helices (TMH) (Lin *et al*., 2012; Kellosalo *et al*., 2012; Li *et al*., 2016; Tsai *et al*., 2019; Vidilaseris *et al*., 2019). A single subunit is formed by two concentric rings of TMH: the inner ring (TMH 5, 6, 11, 12, 15 and 16) makes up the 4 catalytic regions: the catalytic centre, the coupling funnel, the ionic gate and the exit channel, while the outer ring (TMH 1, 2, 3, 4, 7, 8, 9, 10, 13, 14) forms the subunit-subunit interface and the membrane-facing surface of the protein. Despite their common structure, seven different mPPase subfamilies have been functionally characterised (Luoto *et al*., 2011; Tsai *et al*., 2014; Luoto *et al*., 2015). In short, mPPase catalytic activity is either K^+ dependent or K^+ independent and the pumping specificity is either H^+ only, Na^+ only, dual H^+/Na^+ or H^+ regulated by Na^+. Of these subfamilies, only the 3D structures of the K^+-dependent H^+-PPase from *Vigna radiata* (*Vr*-PPase) (Lin *et al*., 2012), and the K^+-dependent Na^+-PPase from *Thermotoga maritima* (*Tm*-PPase) (Kellosalo *et al*., 2012; Li *et al*., 2016; Vidilaseris *et al*., 2019) have been resolved to high resolution.

These 3D structures have facilitated understanding of the mechanism by which mPPases perform hydrolysis and ion pumping. However, there is conflict surrounding the order of these two events (Baykov, 2020; Anashkin *et al*., 2021) and more recently, data have been published indicating that the mechanism is more complicated than previously thought (Artukka *et al*., 2018; Vidilaseris *et al*., 2019). There is also a dispute around how the mPPase subunits

operate with one another, which may be explained by the environment of the protein: lipidated (Artukka *et al*., 2018) or detergent-solubilised (Vidilaseris *et al*., 2019), and so could indicate protein-lipid interactions (Holmes *et al*., 2019). Despite evidence that the lipid environment may have a role in regulating the function of mPPases, the interaction of mPPases with their lipid environment is still unknown.

In recent years, the use of molecular dynamics (MD) simulations has played a role in predicting and subsequently studying membrane protein-lipid interactions (Corradi *et al*., 2019). These interactions can have multiple effects on the protein of interest, for example modulating stability (Habeck *et al*., 2015), assisting conformational changes (Sweadner, 2017), oligomerisation (Gupta *et al*., 2017; Pyle *et al*., 2018) and largescale protein organisation (Kalli and Reithmeier, 2018). MD simulations represent a robust way to identify putative lipid binding sites which can be refined through further simulations or in tandem with other methods (Corradi *et al*., 2019).

In this study, I examined the interactions and dynamics of three mPPases structures in model lipid bilayers *via* multi-scale MD simulations. My results suggest that mPPases form an anionic annulus in the membrane and possess specific anionic lipid binding sites at the dimer interface and the distal regions of the protein. These protein-lipid interactions are conserved between mPPases from different species with differing pumping specificities, which may suggest that these are a general property of mPPases.

5.2 Results

5.2.1 *Tm*-PPase Forms an Anionic Lipid Fingerprint in the Membrane

We first performed simulations with the 3D structure of the *Tm*-PPase due to the extensive structural characterisation of the protein over the last decade (Kellosalo *et al*., 2012; Li *et al*., 2016; Vidilaseris *et al*., 2019). The *Tm*-PPase structure was inserted into bilayers containing POPE and POPA, POPG or POPS molecules. Little is known about the specific compositions of the *T. maritima* native lipid bilayers (Völkl *et al*., 1993; Guan *et al*., 2013; Sohlenkamp and Geiger, 2015) and so palmitoyl-oleoyl phospholipids were considered an appropriate proxy. Following 5 μs of simulation, the normalised contacts of the protein with the anionic component in the inner leaflet of the bilayer were higher for all residues compared with the other lipid components, suggesting that the anionic lipids POPA, POPG and POPS interacted preferentially with *Tm*-PPase in the cytoplasmic leaflet compared to the zwitterionic POPE lipids (**Fig. 5.1**). Titration of the concentration of each anionic lipid, (40%, 20% and 30% (made up of 10% of each anionic lipid) of the overall lipid concentration) demonstrated that this annulus was retained in all concentrations with ~20 lipids interacting with the protein in each

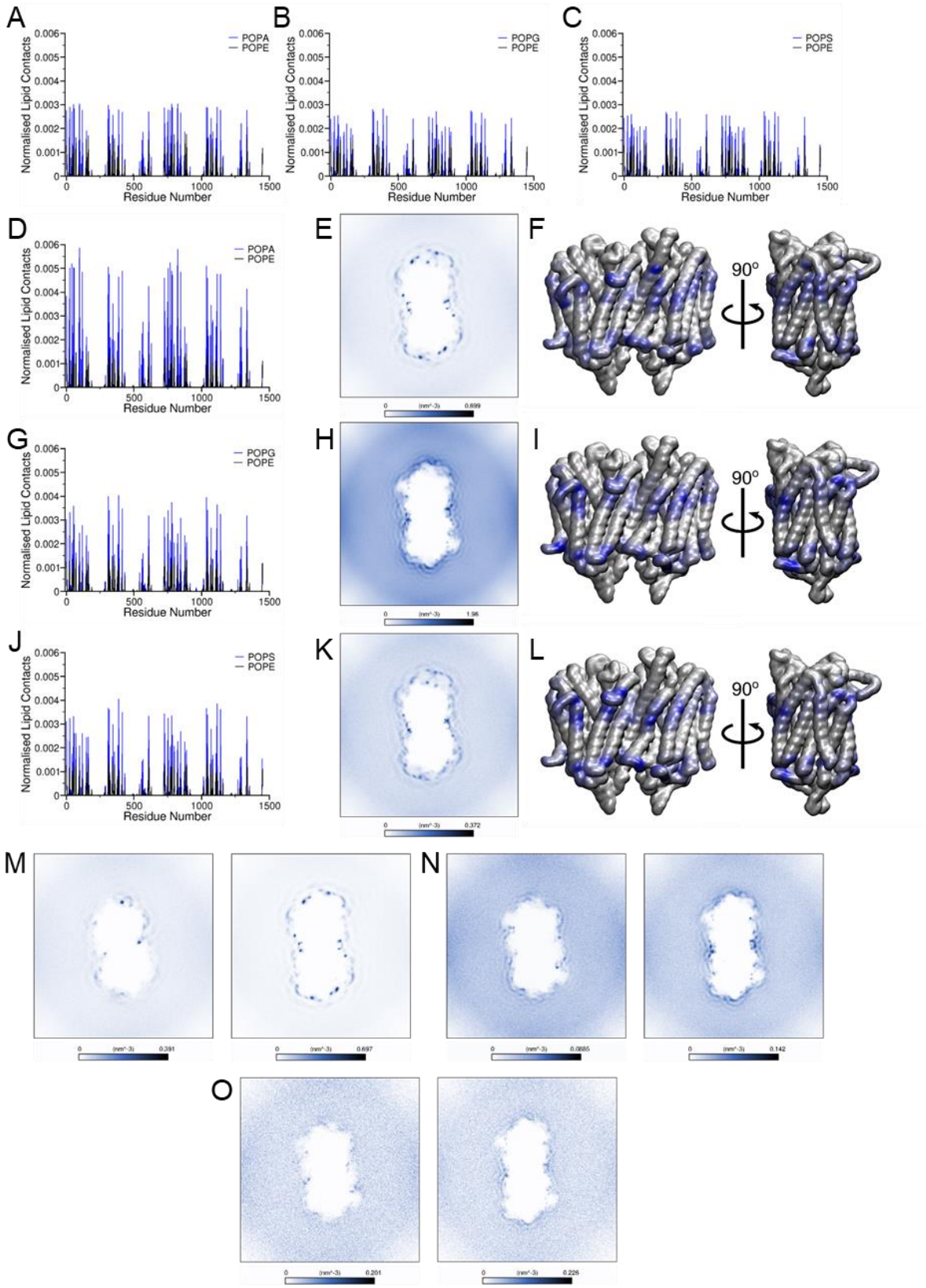


Figure 5.1 *Tm*-PPase Forms an Anionic Membrane Fingerprint During CG Simulations.
The normalised lipid contacts between the backbone particles of *Tm*-PPase and the PO4 particles of the lipids in the **A)** TmPA40, **B)** TmPG40 and **C)** TmPS40 coarse-grained systems following 5 µs of simulation. The normalised lipid contacts alongside the density of the anionic lipid PO4 particles and the contacts plotted on the 3D protein structure of the **D-F)** TmPA20, **G-I)** TmPG20 and **J-L)** TmPS20 systems following 5 µs of simulation. The density maps of each membrane leaflet (left: luminal, right: cytoplasmic) for the **M)** TmPA20, **N)** TmPG20 and **O)** TmPS20 systems.

concentration. Therefore, the 20% anionic lipid bilayers were chosen for further experiments, so selective interactions could be distinguished.

Analysis of the interactions between the *Tm*-PPase and the anionic lipids allowed me to identify lipid binding sites through contact analysis and lipid density (**Fig. 5.1**). These binding sites were identified using the top 5% of interacting residues from the contact analysis. This analysis revealed four symmetrical anionic lipid binding sites at the dimer interface and the "distal regions" of the protein. Nine positive lysine and arginine residues ($R^{1.60}$ (27), $K^{1.61}$ (28), $R^{2.38}$ (43), $K^{3.59}$ (96), $R^{3.63}$ (100), $R^{3.67}$ (104), $K^{4.40}$ (120), $K^{8.40}$ (311) and $K^{8.41}$ (312)) formed each of the distal interaction sites. The residues involved in the dimer interface sites were four positive lysine residues ($K^{9.70}$ (389), $K^{10.49}$ (415), $K^{13.52}$ (568) and $K^{14.48}$ (609)), two of which were located in one subunit and the other two in the other subunit.

Analysis of the lipid contacts revealed that POPA interacted more frequently with *Tm*-PPase compared to POPG and POPS at 20% anionic lipid bilayer content. To examine further the more frequent binding of POPA lipids, I have also performed a simulation in which the bilayer contained an equal anionic lipid mix (10% each of POPA, POPG and POPS). In this simulation, POPA interacted up to 43-fold or 84-fold more than POPG or POPS, respectively, supporting my previous observation that POPA interacts more with *Tm*-PPase compared to other anionic lipids. Therefore, the further characterisation of the *Tm*-PPase lipid binding sites in this work was performed using the system that contained *Tm*-PPase with 20 % POPA lipids in the bilayer (TmPA20: See Methods for system naming conventions) system, but it was found that, for all of the parameters assessed, the interactions with POPG and POPS were reduced compared to those with POPA.

Contact analysis and PyLipID (Song *et al*., 2022b) were used to understand the lipid occupancy, exchange and residence of the distal and interfacial sites (**Fig. 5.2**). The distal site was able to accommodate more lipids, with 2 lipids bound the majority of the time, rarely increasing to 3 or in some cases 4 for short periods. Residues $R^{2.38}$ (43), $R^{3.63}$ (100) and $K^{8.40}$ (311) had the highest lipid residence times (1.41 μs, 0.23 μs and 0.13 μs, respectively) and were found to primarily coordinate the 2 or 3 lipids in the site, with the 4th lipid briefly interacting with other residues in the site. The interfacial site was able to accommodate fewer lipids at any time, usually being occupied by one or occasionally two lipids. Rarely a third lipid would interact with the site, but this was with only one of the outer residues (typically $K^{14.48}$ (609)), rather than being coordinated within the site. This is consistent with the size of the sites, as the interfacial site is much smaller than the distal site, and is situated in an interfacial cleft, causing the inner residues to be less accessible for contacts. Additionally, the rate of lipid

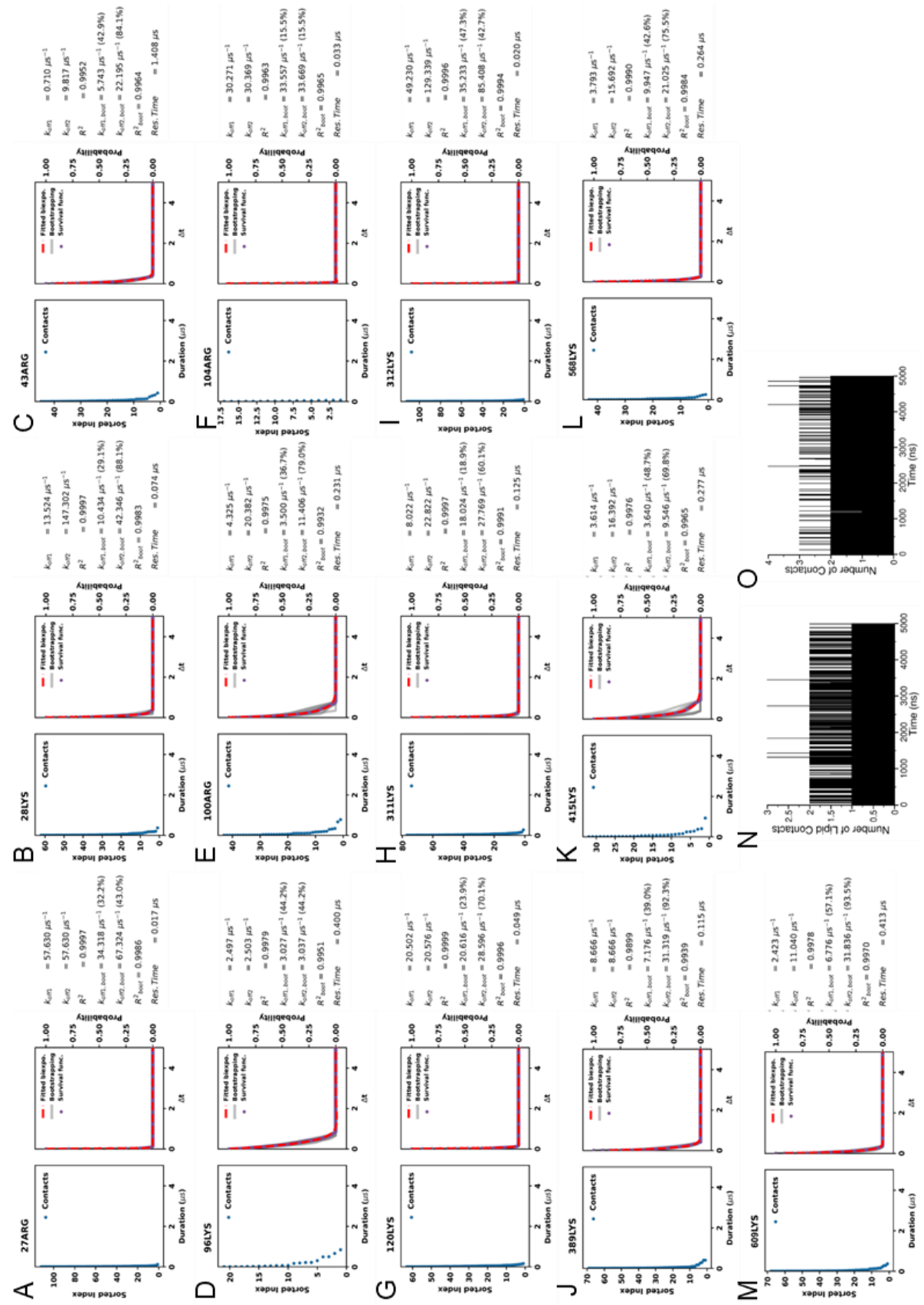


Figure 5.2 The Lipid Residence Time at *Tm*-PPase Residues.
The lipid residence plots and details generated through PyLipID for the residues of the **A-I)** distal interaction site and the **J-M)** interfacial interaction site in the TmPA20 systems. The number of contacts between PO4 particles of POPA with the **N)** interfacial and **O)** distal interaction sites.

exchange was higher in the interfacial site than the distal site, with residence times averaging 0.28 μs for the four interfacial residues.

5.2.2 *Vr*-PPase in its Native Bilayer Forms Similar Membrane Interactions

To examine whether the anionic lipid fingerprint identified above was *Tm*-PPase specific or if it also occurs in mPPases in other species and in other membranes, I also performed simulations using the crystal structure of *Vr*-PPase. Like *Tm*-PPase, *Vr*-PPase is structurally well-characterised but mesophilic rather than thermophilic and as there are no resting state *Vr*-PPase structures available, the relaxed product bound state (PDB: 5GPJ (3.5 Å)) was chosen for simulation to be the most comparable to the *Tm*-PPase findings.

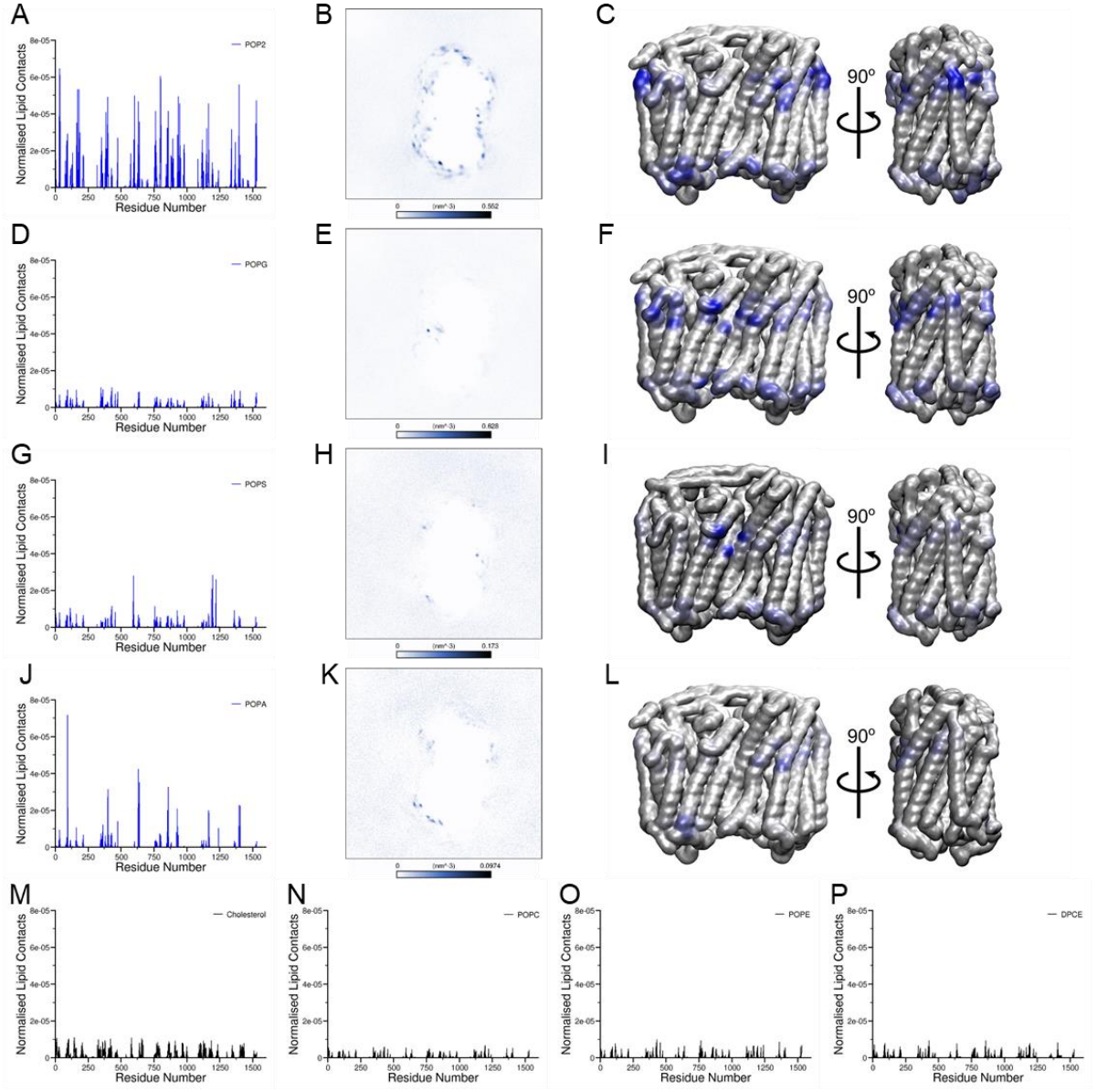


Figure 5.3 *Vr*-PPase in its Native Bilayer Forms an Anionic Membrane Fingerprint.
The interactions between *Vr*-PPase and the PO4 particles of the anionic lipids in a realistic tonoplast membrane represented through normalised lipid contacts, lipid density and the contacts on the 3D protein surface for **A-C)** PIP_2, **D-F)** POPG, **G-I)** POPS and **J-L)** POPA following 5 μs of CG simulation. The normalised lipid contacts with **M)** the entire cholesterol molecule, **N)** the PO4 particle of POPC, **O)** the PO4 particle of POPE and **P)** the PO4 particle of DPCE.

As there is more information about the *V. radiata* tonoplast membrane, simulations of *Vr*-PPase were performed in bilayers resembling this membrane (29% cholesterol, 25% POPC, 17% POPE, 17% ceramide hexoside, 6% PIP_2, 3% POPG, 2% POPS and 1% POPA) (Yoshida and Uemura, 1986). The four anionic interaction sites seen with *Tm*-PPase were present at the interfacial and distal regions of the *Vr*-PPase protein, where lysine and arginine residues composed the distal site ($K^{1.60}$ (35), $K^{1.62}$ (37), $K^{1.67}$ (42), $K^{2.50}$ (94), $K^{3.59}$ (162), $R^{4.33}$ (177), $K^{4.37}$ (181), $R^{4.44}$ (188), $K^{8.41}$ (352), $K^{8.52}$ (363)) (**Fig. 5.3**). However, unlike *Tm*-PPase, the distribution of residues forming the interfacial site was uneven, as one from one subunit and three from the other contributed to forming interactions with the anionic lipids ($K^{10.49}$ (457), $K^{13.48}$ (595), $K^{14.40}$ (632), $K^{14.48}$ (640) for *Vr*-PPase). These simulations were run with 323 K temperature as in my previous simulations, but protein-lipid interactions were very similar in the same simulation in which the temperature was 310 K (**Fig. 5.4**).

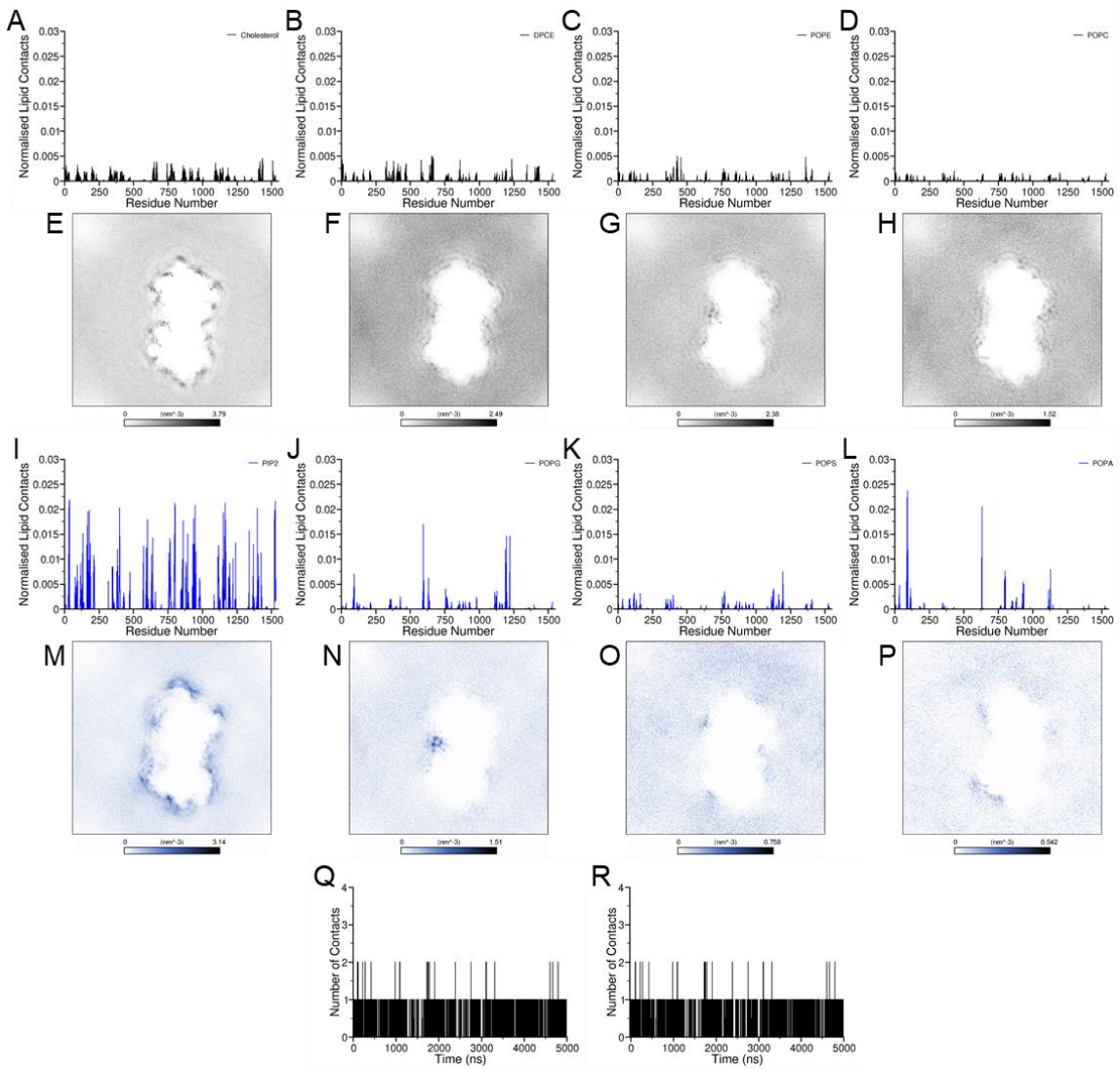


Figure 5.4 The Lipid Contacts with *Vr*-PPase at 310K.
The normalised neutral lipid contacts with *Vr*-PPase at 310 K over 5 µs for **A)** the entire cholesterol molecule, or the PO4 particle of the **B)** POPC, **C)** POPE and **D)** DPCE lipids, and the **E-H)** lipid density of the same groups. The normalised contacts between the PO4 particle of the anionic lipids **I)** PIP_2, **J)** POPG, **K)** POPS and **L)** POPA, and **M-P)** the corresponding lipid density of these lipids. The contacts between the **Q)** interfacial and **R)** distal interaction sites.

As these simulations contained additional anionic lipids compared to the systems with *Tm*-PPase, I was able to study the preference of *Vr*-PPase for anionic lipids in a more realistic membrane. All of the anionic lipids formed interactions with the interfacial and distal interaction sites, but my analysis suggests that interactions between *Vr*-PPase and PIP_2 were highly favoured over the neutral and other anionic lipids. Of the palmitoyl-oleoyl phospholipids, POPA was the favoured lipid over both POPG and POPS, despite its low representation in the membrane (1%). Therefore, the interactions with PIP_2 and POPA were used to understand the properties of the *Vr*-PPase sites.

Despite having similar structural arrangements of the interfacial and distal sites as in *Tm*-PPase, the interactions between the lipids and the protein were somewhat different in *Vr*-PPase (**Fig. 5.5**). The interfacial site appears to be smaller and only accommodates one anionic lipid at a time for the majority of the 20 μs of simulation time, with a second lipid interacting with the binding site very rarely. However, when there are two lipids interacting with the binding side, one lipid interacts with residues $K^{13.48}$ (595) and $K^{14.48}$ (632) that are at the centre of this binding site and the second lipid interacts with one of the outer residues of the site ($K^{10.49}$ (457) and $K^{14.40}$ (640)). Despite this, the average residence time in the interfacial site was longer than that seen for *Tm*-PPase at 0.51 μs. The residence time at each protein residue revealed that there was a lot of variability, and residue $K^{4.37}$ (181) (1.25 μs residence time) had the most prolonged interactions with lipids in this site compared to the others. This uneven residence time may explain why the distal interaction site also appeared to be smaller for *Vr*-PPase compared to *Tm*-PPase, as it formed contacts with a maximum of 3 lipids, but primarily 1 or 2. The reduced occupancy may be due to the increased size of the PIP_2 head group and availability of negative charge to form interactions with, thereby reducing the availability of binding site residues or the tilting of TMH 13 and 14 towards the interface and compressing the site compared to *Tm*-PPase in the resting state (Li *et al*., 2016).

5.2.3 *Pa*-PPase Demonstrates Conservation of the Lipid Fingerprint in Subfamilies

The mPPase from *Pyrobaculum aerophilum* is less well structurally characterised, as only one crystal structure has been solved, to a resolution of 3.8 Å (Strauss, 2021). This structure represents the substrate-analogue bound state, in which the outer TMH are somewhat rearranged in comparison to the *Tm*-PPase and *Vr*-PPase structures used previously in this work. Despite this, identifying whether the anionic lipid fingerprint is conserved across mPPase subfamilies was of interest, and so the structure was prepared for CG simulation, as before.

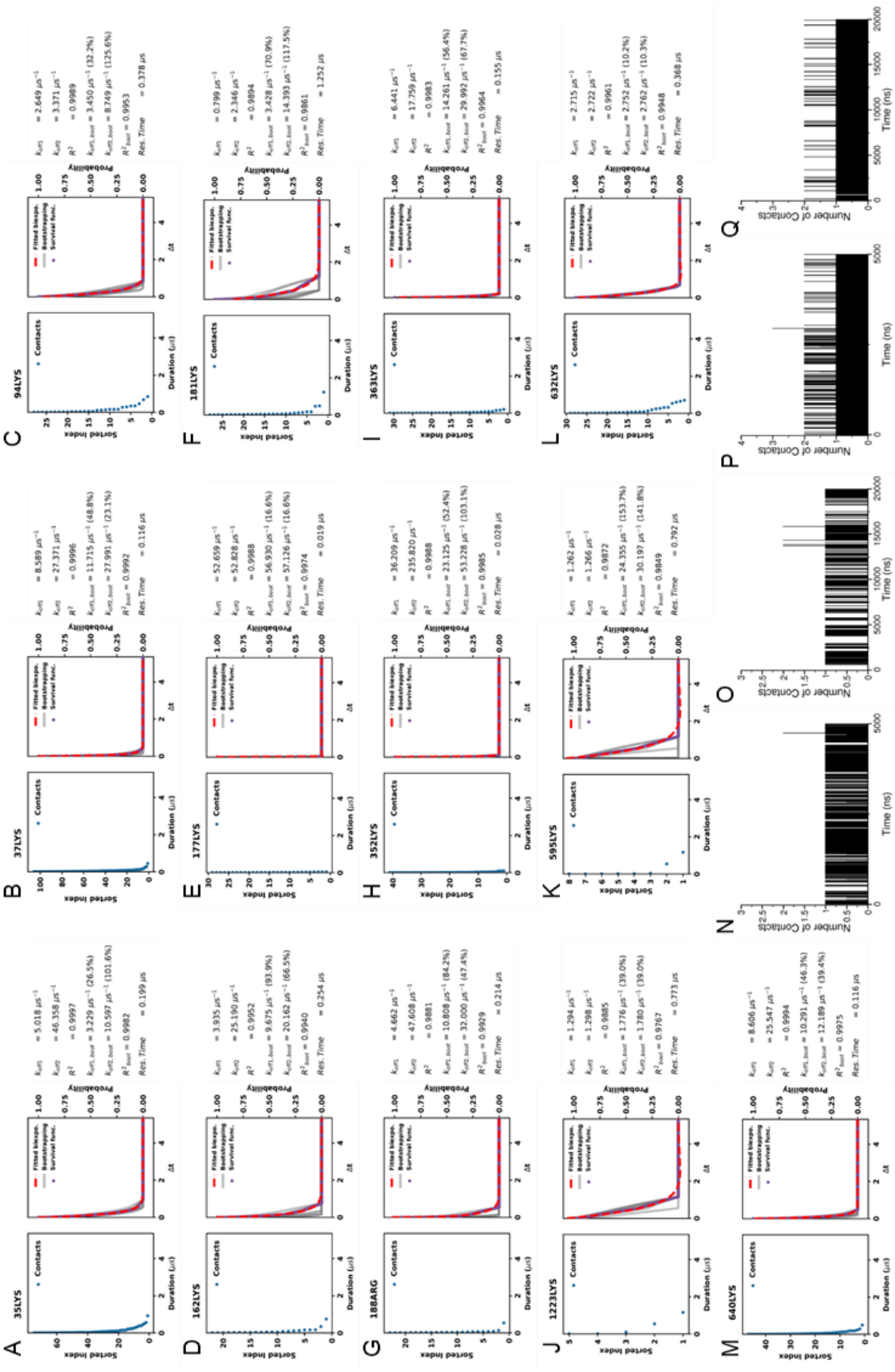


Figure 5.5 The Lipid Residence Time at *Vr*-PPase Residues.
The lipid residence plots and details generated through PyLipID for the residues of the **A-I)** distal interaction site and the **J-M)** interfacial interaction site in the VrTonoplast systems. The number of contacts between PIP_2 and the **N)** interfacial interaction site over 5 μs or **O)** 20 μs and the **P)** distal interaction site over 5 μs or **Q)** 20 μs.

The normalised lipid contacts indicate a marked preference for the anionic lipid species, in particular POPA, as had been seen for the prior mPPases (**Fig. 5.6**). Upon visualisation of the anionic lipid density and the location of the highly interacting residues on the protein structure, the distal and interfacial interaction sites were identified, consisting of $R^{1.60}$ (30), $R^{2.42}$ (50), $R^{2.47}$ (55), $R^{2.50}$ (58), $R^{3.63}$ (109), $R^{4.43}$ (131) and $R^{14.44}$ (593) for the distal site and $K^{9.59}$ (342), $K^{9.73}$ (356), $R^{13.45}$ (549) and $R^{13.52}$ (556) at the interfacial site. This differs significantly from the arrangement of the *Tm*-PPase and *Vr*-PPase interaction sites, as the distal site is smaller, being composed of only 7 residues, including one on TMH 14, and despite the interfacial site being composed of 4 residues, these are from only 2 TMH, rather than 3 or 4. Perhaps due to the sites being smaller, they were able to accommodate less lipids, with a maximum of one POPA lipid binding to the interfacial site at any time, and typically 2 binding to the distal site (**Fig. 5.7**). This may also explain the reduced residence time, as on average this was 0.2 µs for the distal site, ranging from 0.35 µs at residues $R^{1.60}$ (30), $R^{2.47}$ (55) and $R^{3.63}$ (109), and only 0.035 µs at $R^{2.42}$ (50). The interfacial site had an even lower residence time, averaging 0.056 µs, which is over 200 ns less than those in *Tm*-PPase and *Vr*-PPase.

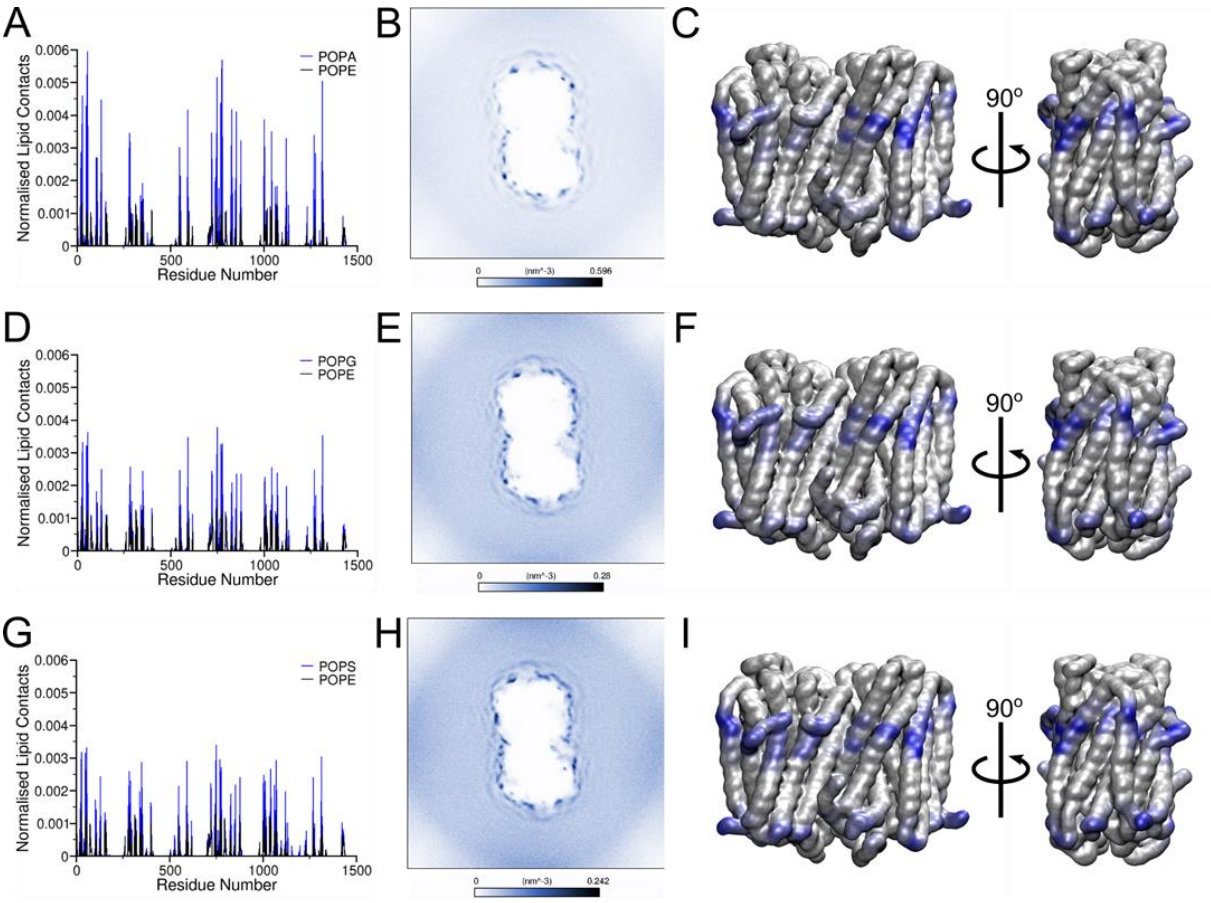


Figure 5.6 *Pa*-PPase Forms an Anionic Membrane Fingerprint.
The normalised lipid contacts alongside the density of the anionic lipid PO4 particles and the contacts plotted on the 3D protein structure of the **D-F)** PaPA20, **G-I)** PaPG20 and **J-L)** PaPS20 systems following 5 µs of simulation.

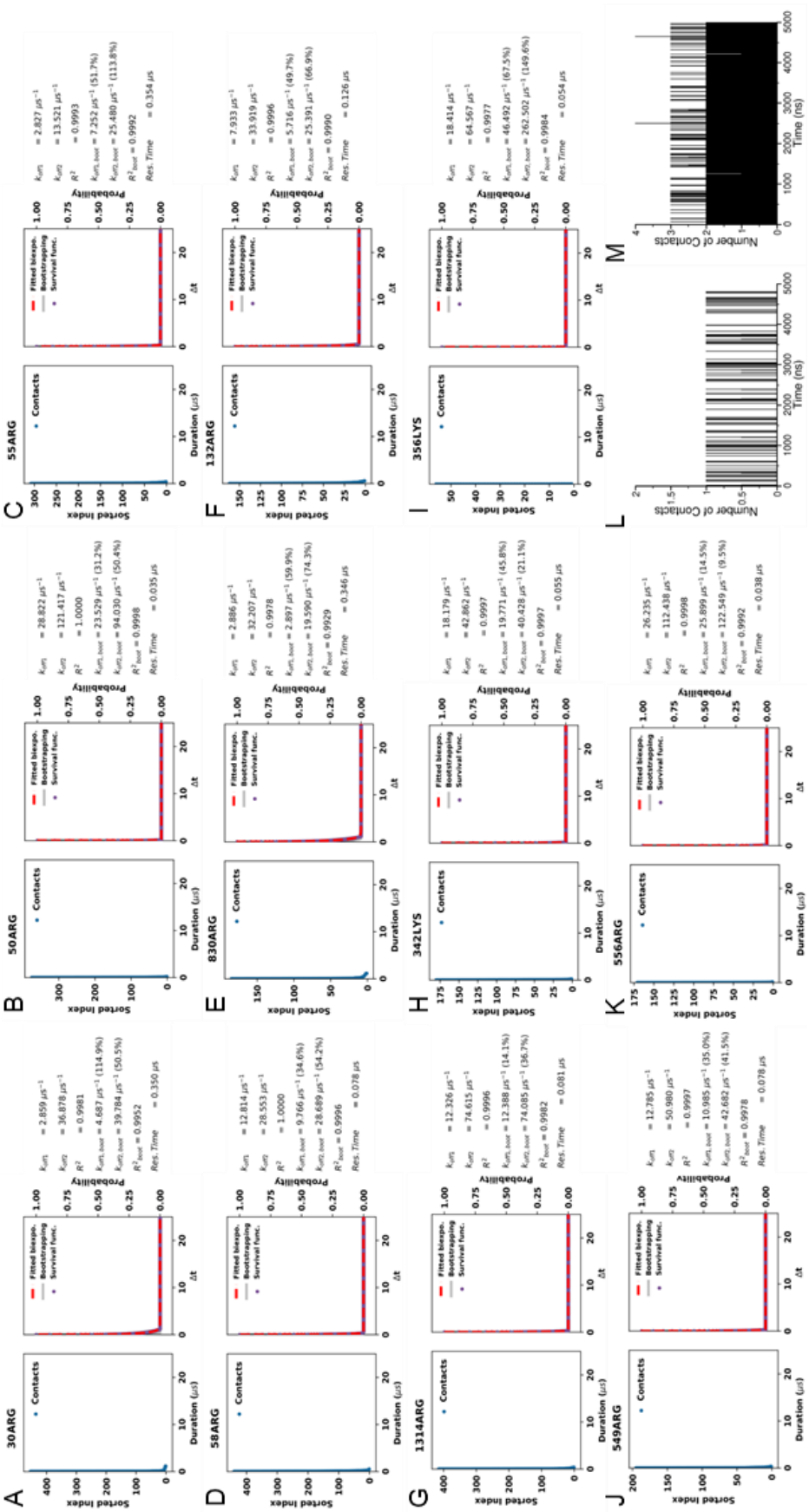


Figure 5.7 The Lipid Residence Time at *Pa*-PPase Residues.
The lipid residence plots and details generated through PyLipID for the residues of the **A-G)** distal interaction site and the **H-K)** interfacial interaction site in the PaPA20 systems. The number of contacts between PO4 particles of POPA with the **L)** interfacial and **M)** distal interaction sites.

This reduced residence time at the interfacial site, alongside the reduced occupancy and the somewhat reduced number of lipid contacts compared to the *Vr*-PPase and *Tm*-PPase structures indicate a reduced lipid preference for the interfacial site. This could be due to either selectivity between mPPase subfamilies (i.e. K^+- dependant or independent), or the catalytic state represented by the *Pa*-PPase structure. Upon binding of substrate, TMH 13 and 14 bend towards the subunit interface, which compacts the available positive residues for anionic lipid interaction. Without further *Pa*-PPase states to simulate, it is difficult to understand whether this effect is state or subfamily related.

5.2.4 *Cp*-PPase-Lipid Interactions Demonstrate the mPPase Fingerprint may be Predictable

Our results above suggest the three structurally characterised mPPases induce an anionic lipid fingerprint in model membranes. As other mPPase studies have also indicated that lipid binding could be of interest for functional and stabilising studies (Artukka *et al*., 2018; Cecchetti *et al*., 2021), I modelled the Na^+/H^+-PPase from *C. leptum* (*Cp*-PPase; see Methods for details) to examine whether the anionic lipid fingerprint is found in other members of this family. The *Cp*-PPase protein structure has previously been modelled for identification of potential stabilising mutations (Harborne *et al*., 2020), but upon inspection of the quality of these models they were rejected for simulation in this work. Despite producing two high quality models from both Robetta and AlphaFold2 with significant similarity (1.05 Å RMSD), the Robetta homology model (reference PDB ID: 5GPJ) was chosen for this study as the arrangement of the distal helices was more in keeping with those seen in *Tm*-PPase and *Vr*-PPase structures (**Fig. 5.8**).

Clostridium species are predicted to contain POPG, POPE, POPS and cardiolipin, in addition to diradylglycerols. For this reason, I simulated this structure in the same bilayers as *Tm*-PPase, as well as enabling better comparison of my results between systems. The *Cp*-PPase simulations were run with 20% POPA, POPG or POPS or an equal anionic lipid mix (10% each). As well as being crystallographic target, *Cp*-PPase is also a different subtype of mPPase compared to *Tm*-PPase and *Vr*-PPase; this allowed me to examine whether the anionic interactions identified for two different mPPases are retained in other subfamilies.

As in the other systems, *Cp*-PPase formed an anionic lipid fingerprint in the membrane, where specific interactions formed between the lipids and positively charged residues at the distal site ($R^{1.56}$ (23), $K^{1.60}$ (27), $K^{2.46}$ (51), $R^{2.47}$ (52), $K^{3.59}$ (101), $R^{7.62}$ (290), $K^{7.64}$ (292)) and at the interfacial site ($K^{9.69}$ (361), $K^{9.73}$ (365), $R^{13.52}$ (543), $R^{14.48}$ (584)) (**Fig. 5.9**). The interfacial site was made up of two residues from each protein subunit and *Cp*-PPase displayed preference for POPA over the anionic lipids POPG or POPS in the mixed simulations. This supports my

previous hypothesis that binding of anionic lipids to these two types of interaction site may be a general property for all mPPases.

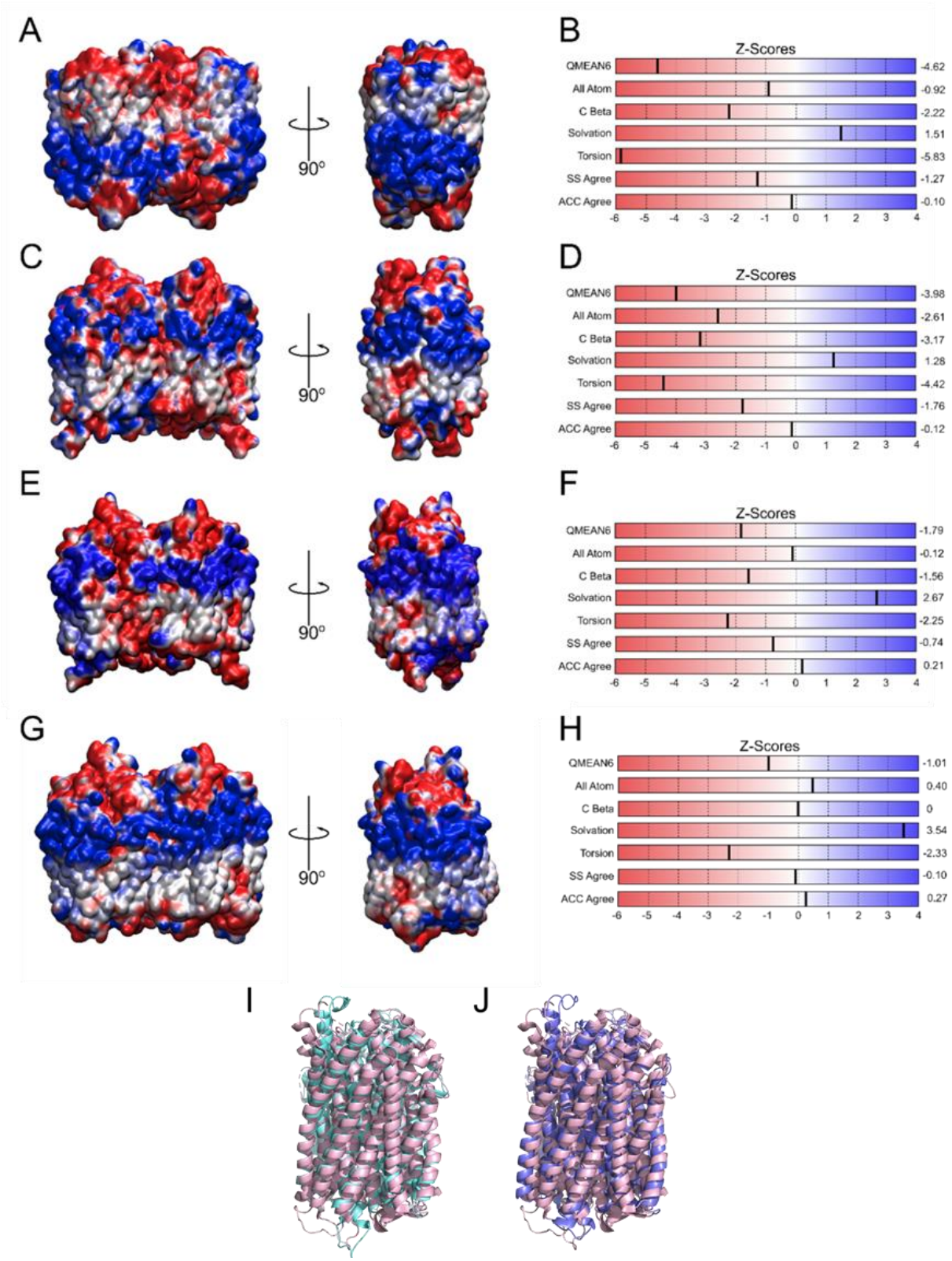


Figure 5.8 The Homology Models of *Cp*-PPase.
The electrostatic profile and Z-Scores of the homology models of *Cp*-PPase generated by **A-B)** iTasser, **C-D)** Modeller, **E-F)** Robetta and **G-H)** AlphaFold2. The comparison of **I)** Robetta (cyan) model and the **J)** Alphafold2 model (purple) with the *Vr*-PPase structure (light pink) (PDB: 5GPJ)
.

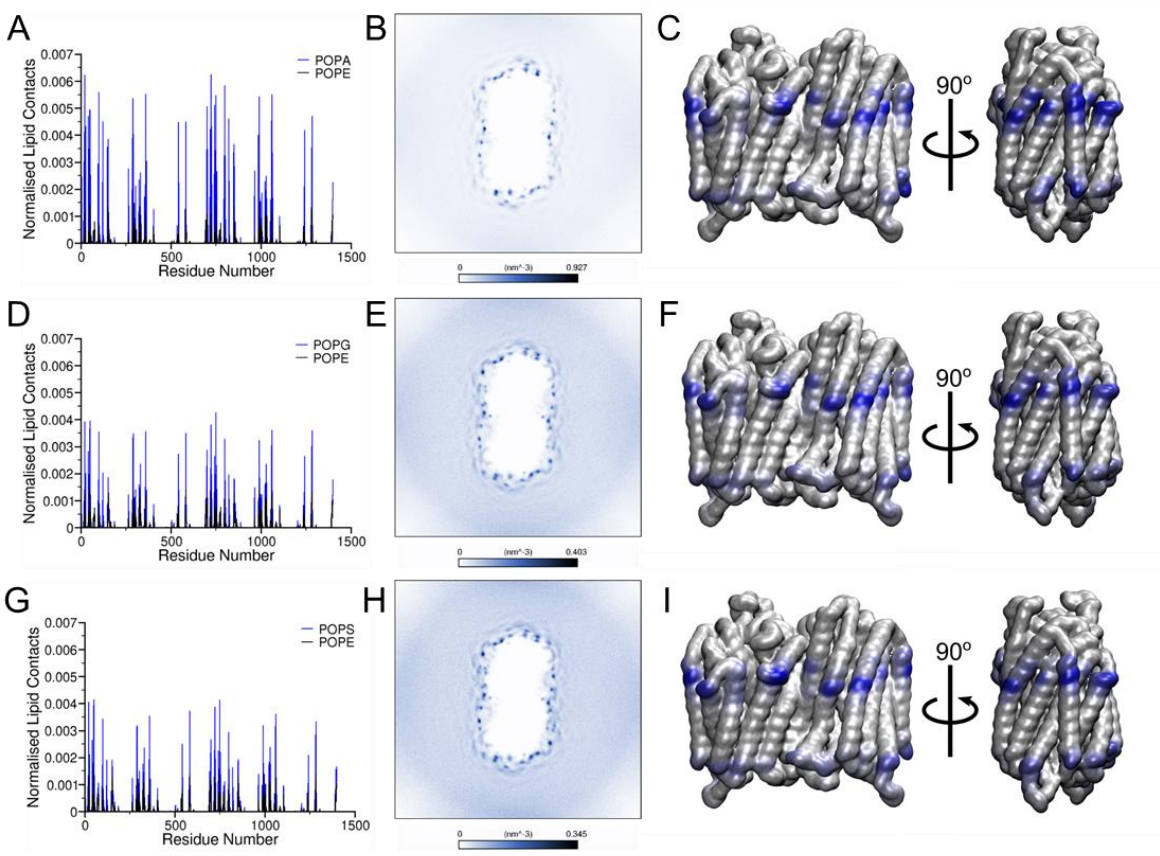


Figure 5.9 Homology Modelled *Cp*-PPase Forms an Anionic Membrane Fingerprint.
The normalised lipid contacts alongside the density of the anionic lipid PO4 particles and the contacts plotted on the 3D protein structure of the **D-F)** CpPA20, **G-I)** CpPG20 and **J-L)** CpPS20 systems following 5 μs of simulation.

Interestingly, the distal and interfacial sites of *Cp*-PPase better resembled that of *Tm*-PPase, despite being modelled with reference to the *Vr*-PPase structure, both in structural arrangement and lipid occupancy and average residence time. The interfacial site accommodates up to 2 anionic lipids at a time and the distal site could accommodate up to 4, despite being comprised of only 7 positive residues, similar to the interactions seen with *Tm*-PPase in the same bilayer systems (**Fig. 5.10**). However, the sites have shorter average residence times than in the other mPPase simulations, with only 0.1 μs for the distal site, primarily contributed to by $K^{2.46}$ (51), $R^{2.47}$ (52), $K^{3.59}$ (101), $R^{7.62}$ (290), and the interfacial site had an average residence time of 0.2 μs.

5.2.5 *In silico* Mutations Disrupt the Protein-Lipid Interactions

To examine whether there was any synergy in the binding of lipids in the lipid binding sites identified above, I performed *in-silico* mutagenesis of the four positively charged residues in the interfacial site to alanine. Analysis of the lipid density following identical simulation to the wild-type (WT) proteins showed that when a single or double site mutation (DSM) was made, the anionic lipid binding was lost, but binding at the intact site remained (**Fig. 5.11**). Therefore, indicating that my simulations do not demonstrate any cooperativity between the binding sites.

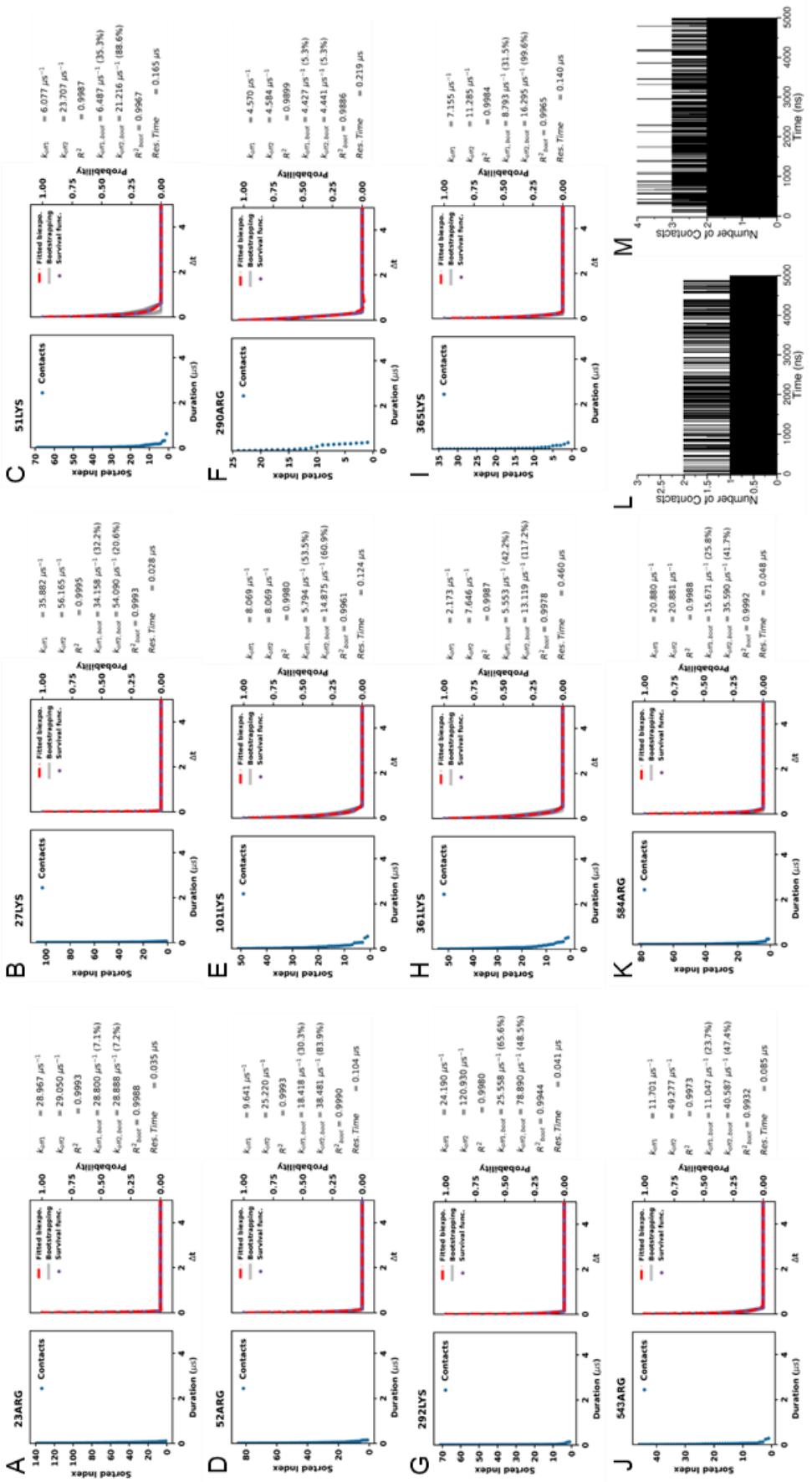


Figure 5.10 The Lipid Residence Time at *Cp*-PPase Residues.
The lipid residence plots and details generated through PyLipID for the residues of the **A-G)** distal interaction site and the **H-K)** interfacial interaction site in the CpPA20 systems.

This confirmed that the anionic lipid binding at these sites was due to specific interactions with the protein, rather than a stochastic effect, and that the binding is independent at each site. To assess whether this was due to the properties of the protein surface in these areas, electrostatic analysis of the proteins was performed and revealed that positively charged regions matched the location of the anionic binding at the interfacial and distal regions of the protein. The electrostatic profiles were very similar between the three mPPases in this study, as well as the arrangement of their binding site residues.

5.2.6 Refinement of the Lipid Interactions *via* Atomistic Simulations

The effect of lipid binding on the dimeric arrangement or stability of the protein could not be assessed through these coarse-grained (CG) simulations, as the elastic network between the

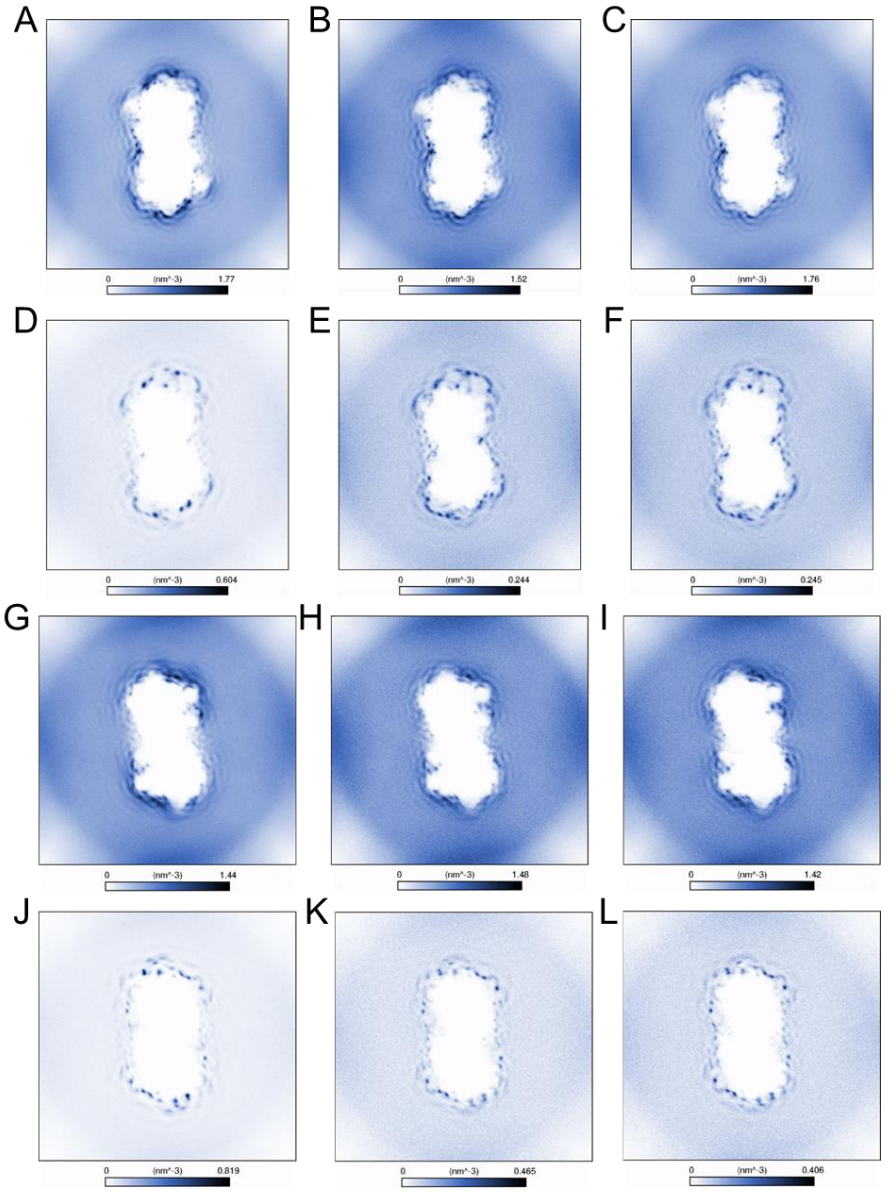


Figure 5.11 Mutagenesis of the Interfacial Site Leads to Loss of Anionic Density.
The density maps depicting the average density of PO4 particles of the anionic lipid in the **A)** TmPA20_SSM, **B)** TmPG20_SSM, **C)** TmPS20_SSM, **D)** TmPA20_DSM, **E)** TmPG20_DSM, **F)** TmPS20_DSM, **G)** PaPA20_DSM, **H)** PaPG20_DSM, **I)** PaPS20_DSM, **J)** CpPA20_DSM, **K)** CpPG20_DSM and **L)** CpPS20_DSM.

protein monomers maintained the protein-protein interface and conformational state, while allowing flexible regions to move. Therefore, to refine the protein-lipid interactions at these sites and identify any effect the lipid binding had on protein structure and dynamics, the final coordinates of the WT or DSM *Tm*-PPase, *Vr*-PPase, *Pa*-PPase and *Cp*-PPase systems were converted to an AT representation. Frames were chosen where at least one anionic lipid was bound to all of the anionic lipid interaction sites so that their binding could be better understood at an atomistic resolution.

The association of the anionic lipids at the WT protein interface was maintained throughout the 250 ns simulations, as was predicted with the CG residence times, and interactions with all the proposed lysine and arginine residues at the mPPase protein interface and distal interaction sites were retained (**Fig. 5.12**)**.** Analysis of the simulations revealed interacting residues that had not been predicted by the initial CG simulations, such as with $Y^{9.64}$ (383),

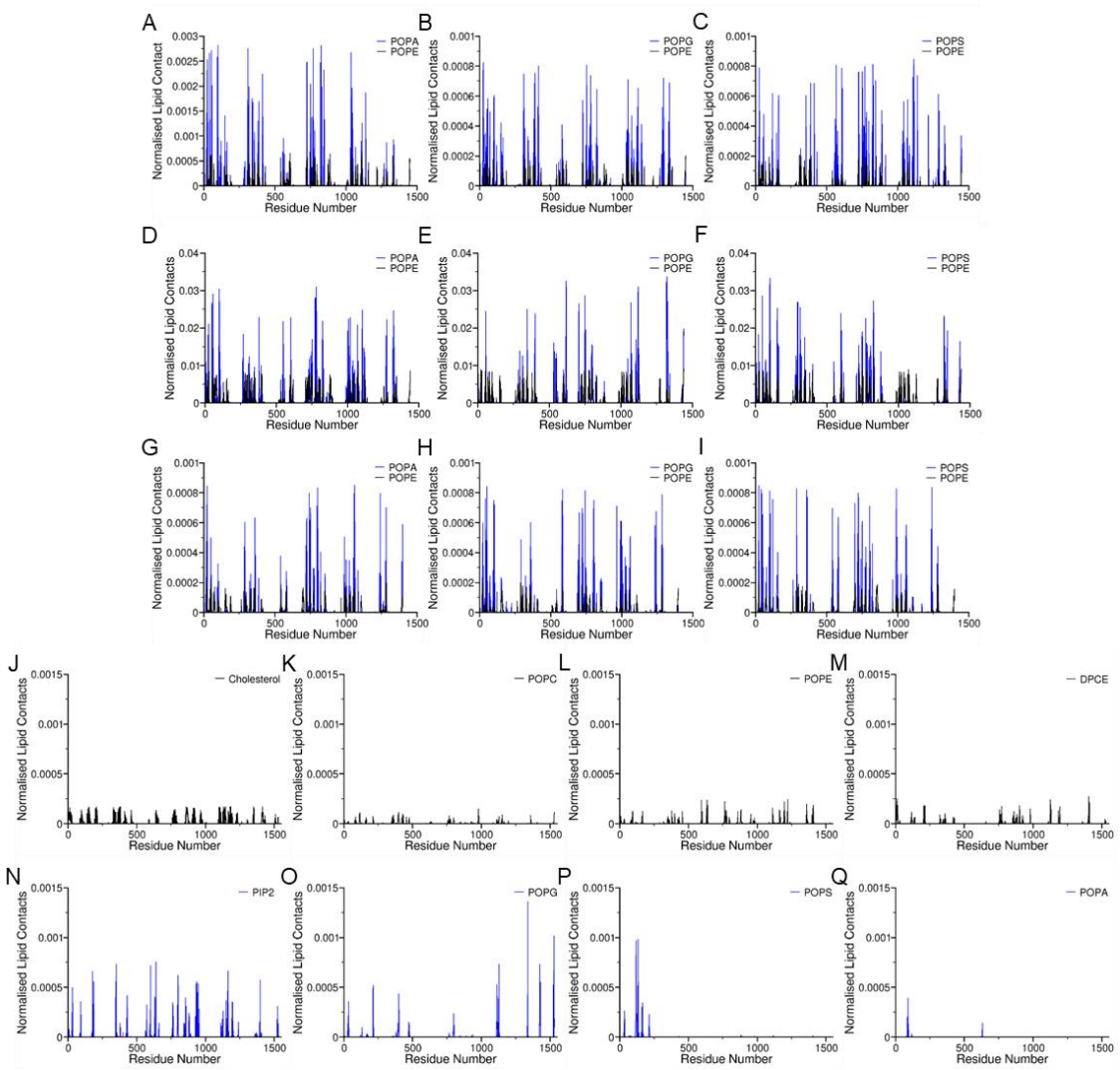


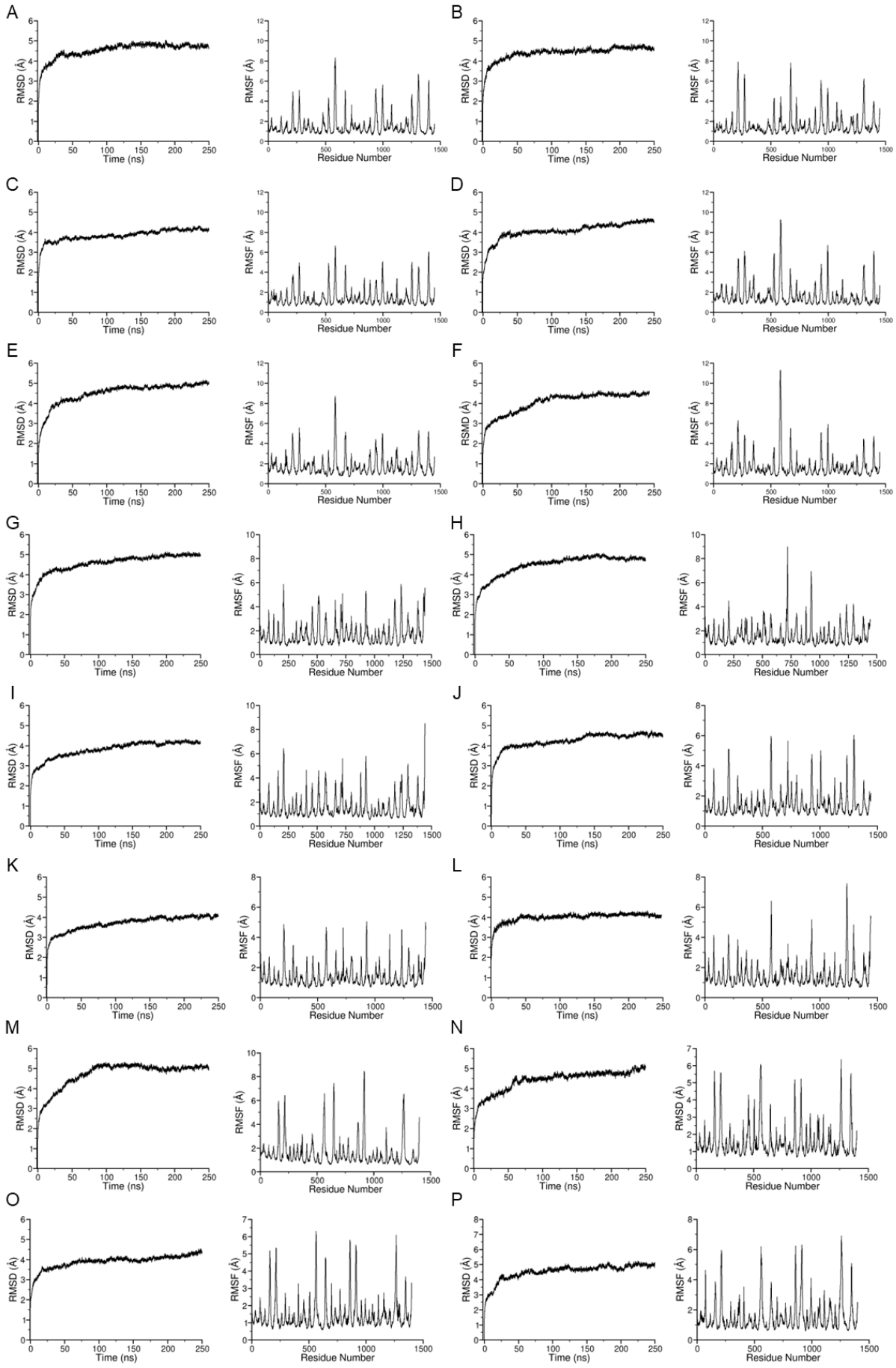
Figure 5.12 Normalised Lipid Contacts in Backmapped Systems Corresponds to the CG Results.
The normalised lipid contacts between the bilayer and lipids in the **A)** TmPA20, **B)** TmPG20, **C)** TmPS20, **D)** PaPA20, **E)** PaPG20, **F)** PaPS20, **G)** CpPA20, **H)** CpPG20 **I)** CpPS20 and **J-Q)** VrTonoplast systems following 250 ns of simulations.

$Y^{13.45}$ (561) and $Y^{13.51}$ (567) for *Tm*-PPase. However, these appear to only interact a proportionately small amount of the time compared to the lysine and arginine residues. In all these simulations there appeared to be no differences between the two interfacial binding sites in each mPPase or in the dynamics between the mutated and WT proteins throughout the atomistic simulations. However, the RMSD of the proteins in the systems TmPA20_DSM, TmPG20_DSM and CpPS20 did not flatten throughout the 250 ns simulations, which may be due to the high flexibility of the 5-6 and 13-14 loops, as seen in the RMSF analysis (**Fig. 5.13**).

5.3 Discussion

In this study, I have created atomistic models of several different mPPases in modelled membranes. My work with *Vr*-PPase represents the first realistic model of an mPPase in a complex bilayer that mimics its native environment. My simulations have shown that mPPases inserted into a bilayer create an anionic lipid fingerprint around them with preferential interactions with anionic lipids at two distinct sites. In mPPases, two pertinent studies of catalytic asymmetry and inter-subunit communication reported different changes in PP_i affinity at the second monomer following binding of substrate at the first. One of the major differences in these experiments was whether the mPPases were in a lipid (Artukka *et al*., 2018) or detergent (Vidilaseris *et al*., 2019) environment. These data suggest that lipid binding could have a modulatory role in inter-subunit communication and catalytic asymmetry, and delipidation through solubilisation and purification procedures thereby influences mPPase function.

In this investigation, mPPases were found to interact preferentially with anionic lipids compared to neutral lipids. This interaction appears to be a result of the charged surface of the protein (**Fig. 5.14**), as demonstrated by the electrostatic profile analysis and loss of interaction following mutation and loss of charge. In addition to this, the pronounced preference for POPA in *Tm*-PPase may be due to its increased negative charge compared to the POPG and POPS lipids (-2 vs -1) or due to its similarities with the *T. maritima* native dietherglycerol phospholipids lipids, as both are anionic and do not possess a headgroup above the phosphate group (Sohlenkamp and Geiger, 2015). This may also be the case for the *Cp*-PPase preferential interactions with POPA. Non-specific lipid interactions have been proposed to affect protein function. For example, cardiolipin binding to UraA (H^+-uracil symporter) (Kalli *et al*., 2015), cytochrome bc_1 (respiratory chain complex) (Arnarez *et al*., 2013) and SecY (proton translocon) (Corey *et al*., 2018) is thought to attract positive ions for pumping. It is unclear whether this functionality exists for other anionic lipids, as it is believed that this proton trap capability comes from the presence of two phosphate groups with different pK_a values (Kates *et al*., 1993). However, more recent studies suggest the pK_a values are similar for each cardiolipin phosphate (Olofsson and Sparr, 2013; Sathappa and Alder, 2016),


A
RMSD (Å)
Time (ns)
RMSF (Å)
Residue Number
B
RMSD (Å)
Time (ns)
RMSF (Å)
Residue Number
C
RMSD (Å)
Time (ns)
RMSF (Å)
Residue Number
D
RMSD (Å)
Time (ns)
RMSF (Å)
Residue Number
E
RMSD (Å)
Time (ns)
RMSF (Å)
Residue Number
F
RSMD (Å)
Time (ns)
RMSF (Å)
Residue Number
G
RMSD (Å)
Time (ns)
RMSF (Å)
Residue Number
H
RMSD (Å)
Time (ns)
RMSF (Å)
Residue Number
I
RMSD (Å)
Time (ns)
RMSF (Å)
Residue Number
J
RMSD (Å)
Time (ns)
RMSF (Å)
Residue Number
K
RMSD (Å)
Time (ns)
RMSF (Å)
Residue Number
L
RMSD (Å)
Time (ns)
RMSF (Å)
Residue Number
M
RMSD (Å)
Time (ns)
RMSF (Å)
Residue Number
N
RMSF (Å)
Time (ns)
RMSD (Å)
Residue Number
O
RMSD (Å)
Time (ns)
RMSF (Å)
Residue Number
P
RMSD (Å)
Time (ns)
RMSF (Å)
Residue Number

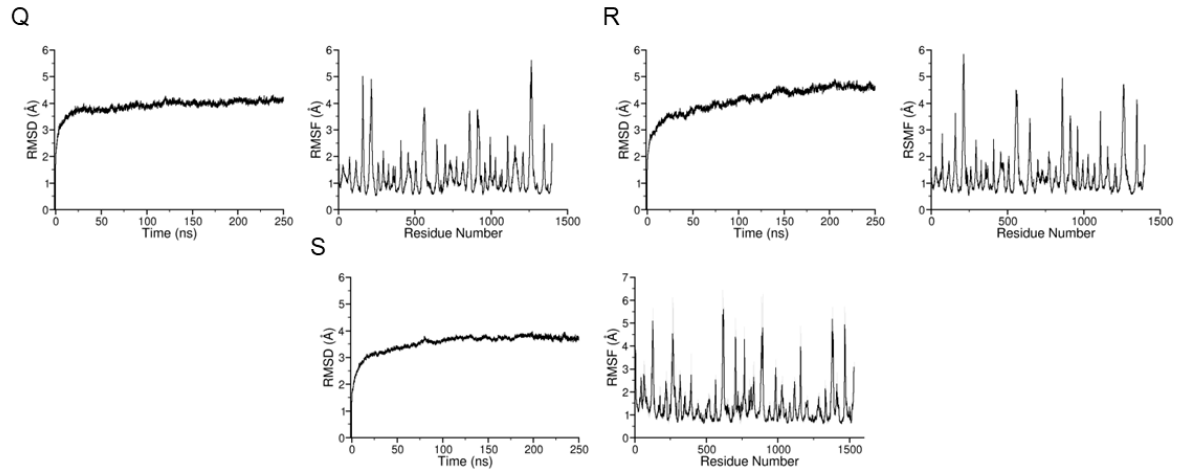


Figure 5.13 The Dynamics of the Backmapped mPPases is Similar Between WT and Interfacial Mutants. The RMSD of the C_α atoms of the protein dimer and the RMSF of the C_α atoms over 250 ns of atomistic simulation after conversion to atomistic representation for the **A)** TmPA20, **B)** TmPG20, **C)** TmPS20, **D)** TmPA20_DSM, **E)** TmPG20_DSM, **F)** TmPS20_DSM, **G)** PaPA20, **H)** PaPG20, **I)** PaPS20, **J)** PaPA20_DSM, **K)** PaPG20_DSM, **L)** PaPS20_DSM, **M)** CpPA20, **N)** CpPG20, **O)** CpPS20, **P)** CpPA20_DSM, **Q)** CpPG20_DSM, **R)** CpPS20_DSM, **S)** VrTonoplast systems.

so potentially this functionality is applicable to the single-phosphate anionic lipids in this investigation. Therefore, the anionic lipid clustering around the mPPases could serve to attract the pumped cations and the Mg_2PP_i for catalysis.

The observed anionic interactions were primarily localised to the cytoplasmic leaflet, with comparatively few contacts with PO4 beads on the opposing side of the protein. Charged regions at the cytoplasmic side of proteins were shown to have a role in membrane protein insertion into the bilayer (Elazar *et al*., 2016). However, the clustering of anionic lipids to these areas and the specific interactions identified in this study may also play role in the function and stability of mPPases. Similar interactions between the lysine and arginine residues of the identified binding sites and the anionic lipid headgroups were also found in other proteins (Haviv *et al*., 2007; Vanegas and Arroyo, 2014; Norimatsu *et al*., 2017), in which the lipid interactions had marked effect on protein stability, function or dynamics. Additionally, it was more recently demonstrated for several oligomeric membrane-integral transporters that interfacial lipids play a crucial role in oligomerisation and stability (Gupta *et al*., 2017; Pyle *et al*., 2018). Lipids are also considered capable of stabilising transient conformational states (Sweadner, 2017). The need to stabilise alternative states of the catalytic cycle and different members of the mPPase family for structural studies means that MD-identified lipid interactions may be a key part of achieving structural information, particularly as interfacial binding sites were identified in this study.

Our studies identified two kinds of lipid binding site on mPPases, an interfacial binding site and a distal binding site. The residues at the interfacial study are not highly conserved, ranging from 47.8% – 61.8% conservation. This is low for functionally relevant residues, but there are

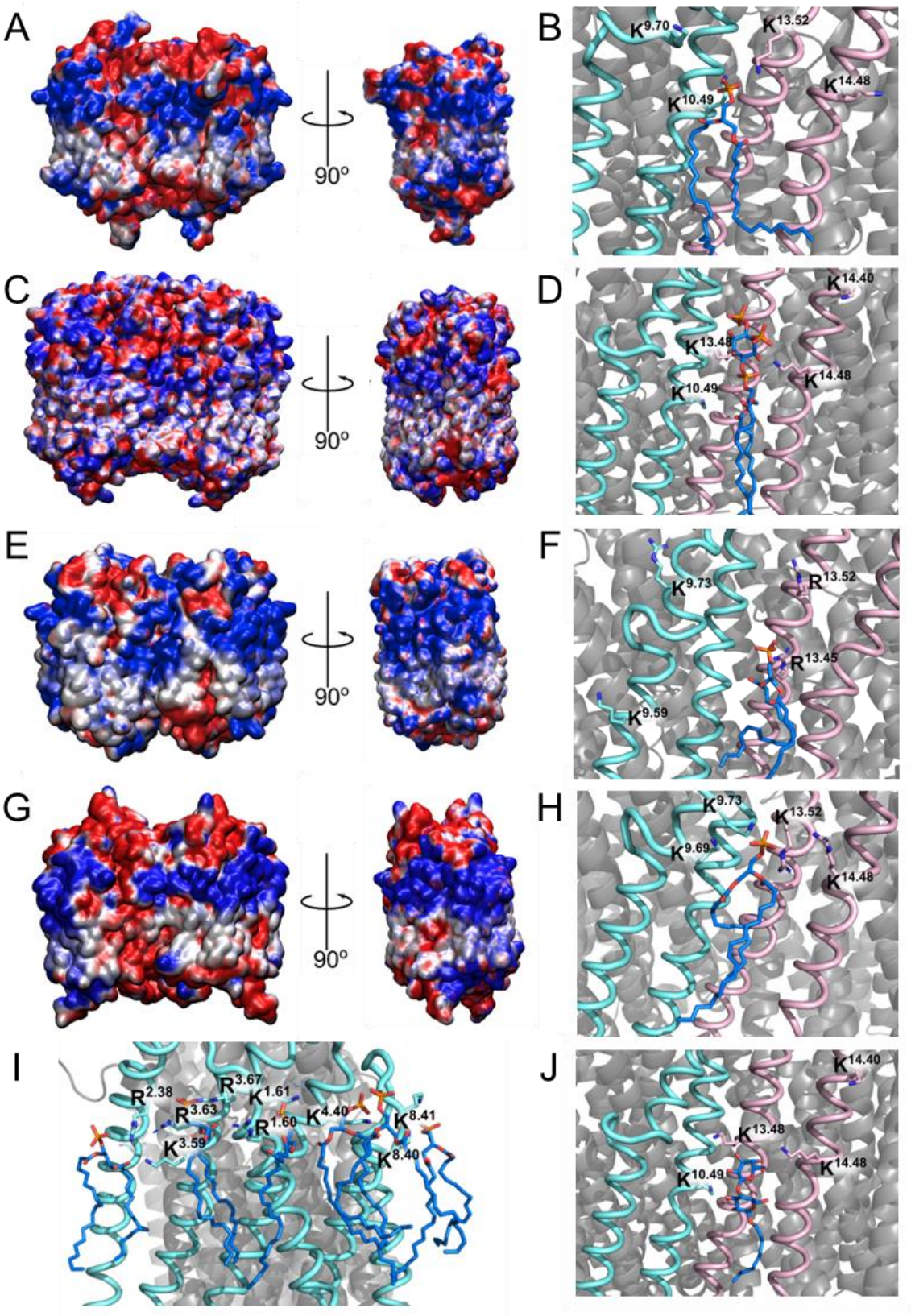


Figure 5.14 Electrostatics profiles and arrangements of the mPPases lipid binding sites.
The electrostatic profile as calculated by APBS and structural arrangement of the **A-B)** *Tm*-PPase, **C-D)** *Vr*-PPase, **E-F)** *Pa*-PPase and **G-H)** *Cp*-PPase protein interfacial anionic lipid binding site with POPA, or in the case of *Vr*-PPase, PIP_2 (blue sticks) bound and the interacting lysines or arginines shown as sticks following 5 µs of simulation. **I)** The distal interaction site of *Tm*-PPase, and **J)** the arrangement of a DM detergent molecule (blue sticks) in the crystal structure of *Vr*-PPase (PDB: 4A01).

the repeated interactions with residues around $K^{9.7X}$, such as $K^{9.70}$ (*Tm*-PPase) and $K^{9.73}$ (*Pa*-PPase and *Cp*-PPase), around K/$R^{13.52}$ (K in *Tm*-PPase, *Vr*-PPase ($K^{13.48}$) and R in *Pa*-PPase ($R^{13.45}$ and $R^{13.52}$) and *Cp*-PPase) and residues around K/$R^{14.48}$ ($K^{14.48}$ in *Tm*-PPase, $K^{14.48}$ in *Vr*-PPase and $R^{14.48}$ in *Cp*-PPase). Many of these differ in position by one helix turn, and so face out into the membrane, thereby preserving their function and potentially accommodating the change in headgroup size between POPA and PIP_2. In addition to conservation of interactions with specific residues, they are positioned on functionally relevant helices. This clustering of interactions on TMH that form the dimer interface and are linked to inter-subunit communication and K^+-dependency clearly supports my hypothesis that presence of anionic lipids at the interfacial interaction site may be functionally relevant to the mPPase catalytic cycle.

In addition to the computational evidence laid out in this work, there is experimental evidence of binding at the interfacial site. In the highest resolution mPPase structure currently available (*Vr*-PPase (PDB: 4A01)) (Lin *et al*., 2012), n-decyl-β-D-maltopyranoside (DM) detergent molecules are situated in this proposed binding site (**Fig. 5.14**). In particular, the binding of one of these is highly similar to the anionic lipid positioning seen in my CG and AT simulations. Moreover, hydrogen bonds form between this detergent molecule and the $K^{9.70}$ and $K^{10.49}$ residues, mirroring the *Tm*-PPase simulations. The structural evidence of binding at the interfacial site in an mPPase provides further support to this region acting as a lipid binding site. The replacement of lipid with detergent was likely due to the purification and crystallisation conditions promoting removal of even tightly bound lipids (Yeagle, 2014). However, the binding of this detergent may have acted similarly to a lipid and helped maintain oligomeric stability, particularly as it has been suggested that detergents can bind in lipid interaction sites (Yeagle, 2014).

The interacting residues at the distal interaction site are not highly conserved (20.8% – 64%), but interactions cluster at similar structural positions. In all proteins in this study, interactions between anionic lipids and the proteins were formed with TMH 1, 2, 3 and 4, with R/$K^{1.60}$ (R in *Tm*-PPase, and K in *Vr*-PPase *Cp*-PPase). Positively charged residues within one helix turn from R/$K^{1.60}$, at $K^{1.61}$ (*Tm*-PPase), $K^{1.62}$ (*Vr*-PPase) and $R^{1.56}$ (*Cp*-PPase) and residues in the midpoints of the helices: $K^{3.59}$, $R^{3.63}$ and $R^{4.48}$ (*Tm*-PPase), $K^{2.50}$, $K^{3.59}$ and $K^{4.44}$ (*Vr*-PPase), and $K^{2.46}$, $R^{2.47}$ and $K^{3.59}$ (*Cp*-PPase) also formed large number of interactions with anionic lipids. In previous simulations of *Tm*-PPase, the distal region of the protein has been found to be flexible (Shah *et al*., 2017), and does not have high conservation between species. Therefore, my simulations suggest that the interactions are conserved to specific helix turns rather than individual residues. This might explain why despite the lower sequence similarity, I observe similar interactions with anionic lipids in the distal area. The distal lipid interaction

sites may play a role in mPPase stability, as the region was identified during a case study of the IMPROvER (integral membrane protein stability selector) pipeline for stabilising mutations (Harborne *et al*., 2020). Further, three of the *Cp*-PPase residues identified by IMPROvER were found to interact with anionic lipids in my study ($F^{1.53}$, $R^{3.67}$ and $R^{7.62}$). I was unable to study effects of lipid interaction on protein stability, as in the CG simulations the elastic force network prevents the protein structure from substantially changing, although the loops were able to move. Conversion to all-atom systems overcame this, as the elastic force network was removed, but no clear stability-based conclusions could be drawn from the dynamics in these simulations. Despite this, due to the corroborative results between this study and the IMPROvER study, the distal lipid interaction sites found in this work may play a role in protein stability.

This study also provides evidence that lipid interactions in the distal and interfacial region may be a more general property of mPPase subfamilies, as *Tm*-PPase is a K^+-dependent Na^+-PPase, *Vr*-PPase is a K^+-dependent H^+-PPase, *Pa*-PPase is a K^+-independent H^+-PPase, and *Cp*-PPase is a K^+-dependent Na^+/H^+-PPase. Despite the conserved pattern of interactions in my studies, there was no conserved sequence motif to identify these interactions in other family members, so likely homology modelling and electrostatic profile analysis, as performed here for *Cp*-PPase, would be required to identify binding sites in other homologues. The retention of interactions between pumping specificities is perhaps not unexpected, as the residues involved in coordinating the pumped ion are at the centre of the mPPase structure (Holmes *et al*., 2019). However, the conservation across subfamilies of lipid interactions that may stabilise the protein and be of mechanistic relevance bodes well for future structural and functional research using alternative homologues.

The preferential interaction of *Vr*-PPase with PIP_2 in this work was striking, as PIP lipids are known to have roles in signalling and protein-protein interactions and are often found bound to proteins of interest (Corradi *et al*., 2018). The function of *Vr*-PPase in the tonoplast membrane of plants has been linked to auxin regulation and signalling (Li *et al*., 2005) and cooperative function with soluble pyrophosphatases (Segami *et al*., 2018); this, taken with the evidence of PIP_2 binding in this study, may suggest a mechanism by which this signalling is able to take place. Additionally, the roles of the vacuolar ATPase complex and *Vr*-PPase are closely aligned (Holmes *et al*., 2019), which may also be mediated by PIP_2 binding and activity.

5.4 Conclusion

This work provides the first evidence that interactions can form between mPPases and anionic lipids and that these are in regions that may hold functional significance and are conserved across homologues. These observations are very promising for future mPPase research. The

role of the distal interaction site in stability needs to be investigated further, but co-crystallisation with lipids or mutagenesis to stabilise this region in detergent may be promising for structural studies and characterisation of other mPPases. Additionally, the putative role of the interfacial site in inter-subunit communication and catalysis warrants further investigation as it may be another factor in the apparent increasing complexity of the mPPase catalytic cycle.

6. Atomistic Simulation Investigations of the mPPase Catalytic Cycle

6.1 Introduction and Rationale

Since 2012, structures of *Tm*-PPase (Kellosalo *et al*., 2012; Li *et al*., 2016; Vidilaseris *et al*., 2019) and *Vr*-PPase (Lin *et al*., 2012; Li *et al*., 2016; Tsai *et al*., 2019) have become available through x-ray crystallography efforts, which has allowed study of conformation changes between sites and theories around the mechanism of ion binding and pumping (Baykov *et al*., 2013; Kajander *et al*., 2013), catalysis (Li *et al*., 2016; Shah *et al*., 2017) and intersubunit communication and cooperation (Artukka *et al*., 2018; Vidilaseris *et al*., 2019). Despite the wealth of structures available, static models of proteins can be difficult to interpret and use in understanding the catalytic cycle, which has led to questions surrounding the molecular level details of the cycle.

The catalytic cycle has been described in the literature as shown in **Figure 6.1**. The resting state (PDB: 4AV3) (Kellosalo *et al*., 2012) transitions following binding of Mg_2PP_i to its active

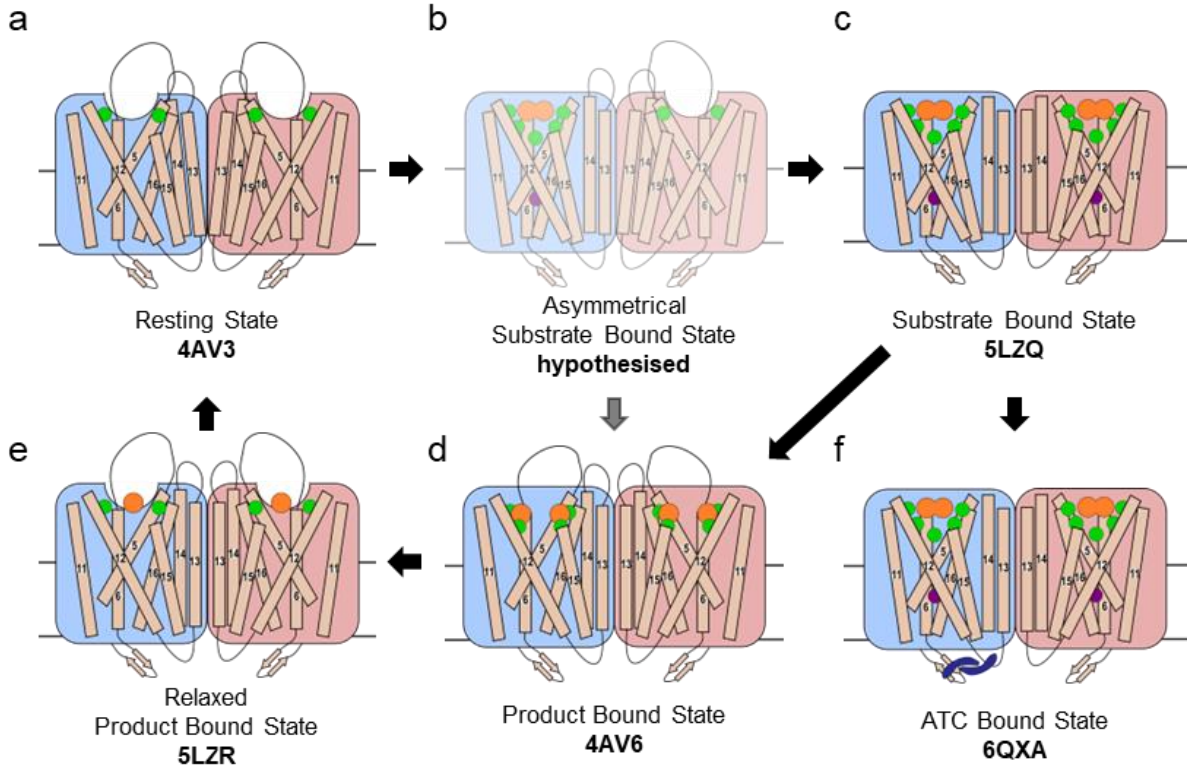


Figure 6.1 Our current understanding of the mPPase catalytic cycle is through crystal structures of *Tm*-PPase
The cycle consists of the resting state (a), that transitions to substrate-bound states that have been hypothesised to either be asymmetrical (b) and catalytically active, or symmetrical with reduced catalytic activity (c). Structurally, this conformational change consists of occlusion of the active site by the 5-6 loop and constriction of the 6 inner TMH 5, 6, 11, 12, 15 and 16, as well as bending of outer TMH 13 and 14 towards the dimer interface. Hydrolysis and ion pumping occur, although the order or mechanics of this have not been fully structurally defined. Following hydrolysis there are two product bound states that are defined by the presence of two (d), then one phosphate and the relaxation of the 5-6 and 13-14 loops (e). Another state that has been solved is the inhibited asymmetrical ATC_2-bound state (f), that suggests binding at the bottom of the protein can inhibit change towards the cytoplasmic side. Mg^{2+}: green spheres, Na^+: purple spheres, PP_i and PO_4: orange spheres, ATC: blue dimer.

site, which already contains 2 magnesium ions and a cation bound at the ionic gate, in the case of ion-pumping mPPases. This substrate binding causes closure of the 5-6 loop over the hydrolytic centre, which may be stabilised by the ordering of the 13-14 loop over the 5-6 loop. The active site rearranges by movement of TMH 11, 12, 15 and 16, thereby constricting the active site and reducing its volume, and the whole dimeric structure shifts through bending of TMH 13 and 14 towards the interface, as seen in the IDP-bound structure (PDB: 5LZQ) (Li *et al*., 2016). There are then three mechanisms for hydrolysis of the phosphoanhydride bond of the PP_i and the coupled ion pumping: the "binding change" mechanism, where the ion is pumped upon binding of substrate (Kellosalo *et al*., 2012), the Grotthus type mechanism (Lin *et al*., 2012), where hydrolysis triggers release of ions at the ionic gate, or the "direct coupling" mechanism, where the proton generated through hydrolysis is the one pumped (Baykov *et al*., 2013). Despite the uncertainty, structurally it has been indicated that TMH 12 potentially moves downwards by 2 Å, which may cause deprotonation of $D^{6.43}$ and $D^{16.39}$, which are then able to activate the nucleophilic water and trigger hydrolysis, although the resolution of these structures impedes clear conclusions about the motion of TMH 12. The protein then adopts the product bound state (PDB: 4AV6) (Kellosalo *et al*., 2012), and the resulting PO_4 leave individually, causing transition to the relaxed product bound state (PDB: 5LZR) (Li *et al*., 2016) before fully relaxing to the resting state.

However, asymmetry in catalysis and the ion position or pumping mechanism has not been fully elucidated structurally. Therefore, this work used simulation techniques with structural models, and hypothesised states, generated through the removal and addition of ligands and ions, to gain insight into the molecular and dynamic detail of the mPPase catalytic cycle, as MD simulations are an established method of studying structural transitions at molecular detail (Kalli *et al*., 2015). The findings were then used to demonstrate the usefulness of these insights in further mPPase study, such as inhibitor discovery and design.

In this chapter, the mPPase catalytic cycle was investigated using a number of methods: simulation of published structures, to understand the differences in dynamics between the resting, substrate-bound and product-bound states and investigate the physiological relevance of these; simulation of modified *Tm*-PPase structures, where the contents of the active site was removed or altered (addition of substrate to the resting state, removal of substrate from the substrate-bound state and replacement of substrate analogue in the substrate-bound state), to explore whether transitions between states could be captured through simulation; and simulation of hypothesised asymmetric structures, where the active site of each subunit was modified independently, to gain an asymmetric non-inhibited catalytic state. Following this, the newly generated models were used for structure-based inhibitor identification, to assess their use for continued mPPase investigation.

6.2 Results and Discussion

6.2.1 Protein Dynamics Vary Greatly Between Different States

6.2.1.1 The mPPase Dimer is Highly Dynamic in the Resting State

The resting state of *Tm*-PPase is characterised by an open 5-6 loop over the active site, which contains two dicationic metal ions (Kellosalo *et al*., 2012). Physiologically, these should be Mg^{2+}, and so the Ca^{2+} ion in the crystal structure was replaced with a Mg^{2+} ion for these simulations, particularly as calcium is known to be inhibitory (Kellosalo *et al*., 2012) and may alter the dynamics and findings of this study if included. The mPPase was simulated in a POPE (80%) and POPG (20%) modelled bilayer, in contrast to previous MD simulation studies that have used non-charged bilayer. Following the findings of the previous chapter, including anionic lipids in this study of mPPase behaviour throughout the catalytic cycle was felt to be relevant to their behaviour.

Following 250 ns of simulation time, RMSD/C_α analysis demonstrated an initial ~2.5 Å change in arrangement compared to the crystallographic structure, following this initial change the RMSD/C_α plateaued after 50 ns (**Fig. 6.2**) and was maintained following a further 500 ns of simulation in the case of one replicate. The RMSF/C_α analysis revealed that the 5-6 and 13-14 loops were particularly flexible and the most dynamic regions of the protein, which was seen previously (Shah *et al*., 2017). Due to the implications of asymmetry during the catalytic cycle, each subunit was analysed individually to assess whether there were differences between them. However, none were seen at this stage of the catalytic cycle.

However, small structural rearrangements of the helices did occur during those first 50 ns of simulation, which were then maintained throughout the rest of the simulation time, likely due to equilibrating to a membrane environment. However, the movement of TMH 12 was very striking (**Fig. 6.2**). This helix was previously thought to move upon binding of the substrate Mg_2PP_i, through a 2 Å downward motion linked to the loss of the ion pair $D^{6.43}$-$K^{12.50}$ (Li *et al*., 2016). However, this work shows that the ion pair is not maintained, as the cytoplasmic end of TMH 12 curves away from the active site by 4.1 Å (as measured between the C_α atoms of $K^{12.50}$ pre and post simulation). This demonstrates the power of MD simulations to provide further insight into crystallographically determined structures, as much of the work understanding the catalytic cycle has centred on the motions of TMH 12, which do appear to be functionally important, but understanding of when these movements occur, and physiological starting positions are essential for insightful interpretation.

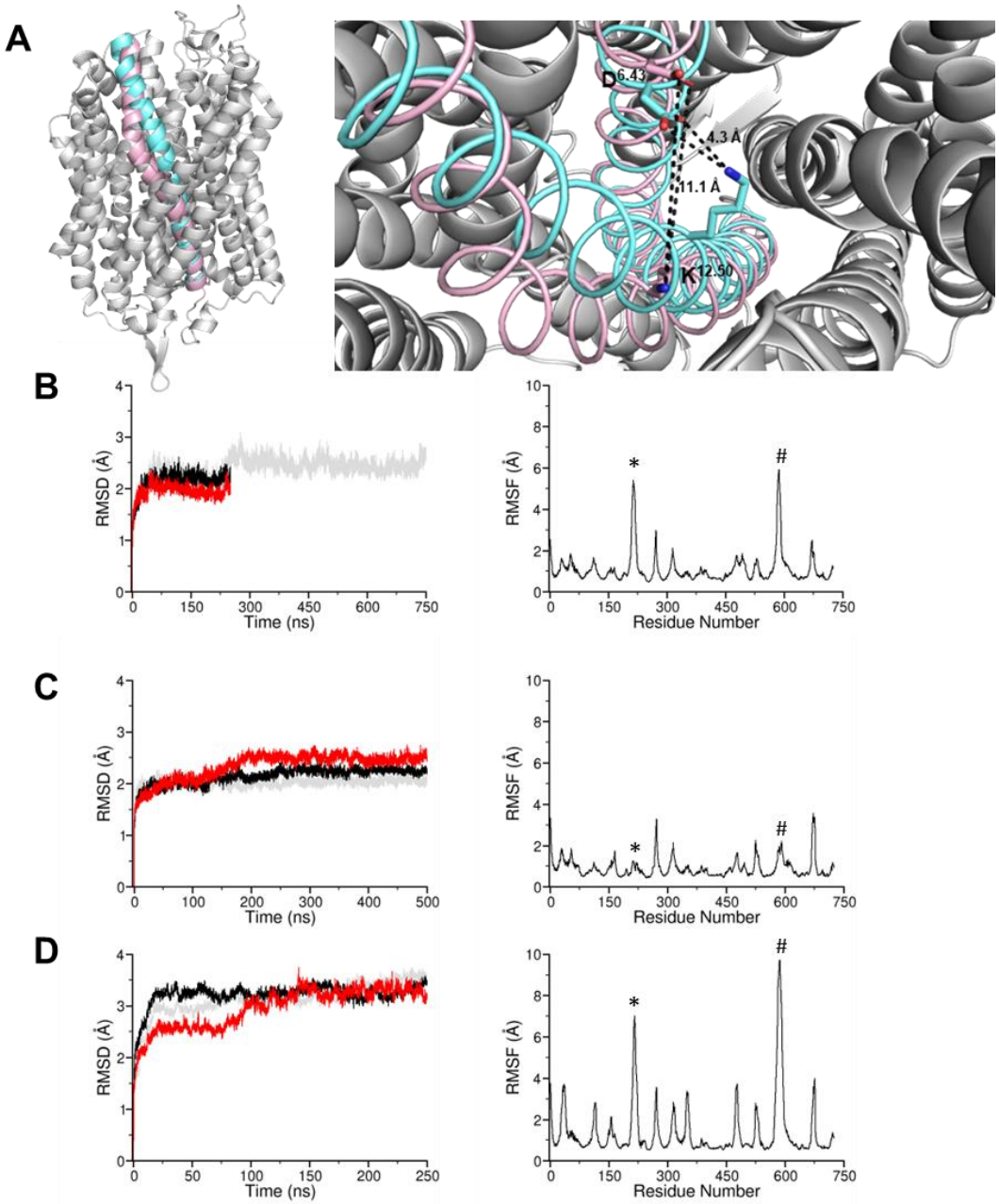


Figure 6.2 Simulations reveal structural changes and dynamic changes between states
A) The resting state adopts a more relaxed structure in simulation, defined by lateral motion of helix 12 away from the active site, as can be seen by the view in the membrane plane (pre-simulation helix in light blue, and post simulation in light pink), likely due to loss of the ion pair between $D^{6.43}$ and $K^{12.50}$, as viewed by the cytoplasmic side with distances between the ends of the side chain displayed. The RMSD and RMSF of the **B)** resting state, **C)** IDP- and **D)** product-bound states, where the 5-6 and 13-14 loops are denoted by * and # respectively. For the RMSD analyses, only the C_α of the helical residues were considered and each replica is shown. For the RMSF, the C_α of each residue was considered and averaged between the replicas (n = 3) and the two subunits.

In addition to the RMS-based dynamics and the structural analyses, distance measurements between residues $T^{5.70}$ and $A^{13.30}$ were performed to complement Pulsed Electron-Electron Double Resonance (PELDOR) experiments performed by Dr Christos Pliotas, which indicated that there are multiple populated conformations of the resting *Tm*-PPase. The six possible distances between these residues were measured over the 250 ns simulations and distance histograms produced. These data were in good agreement with the PELDOR data (**Fig. 6.3**),

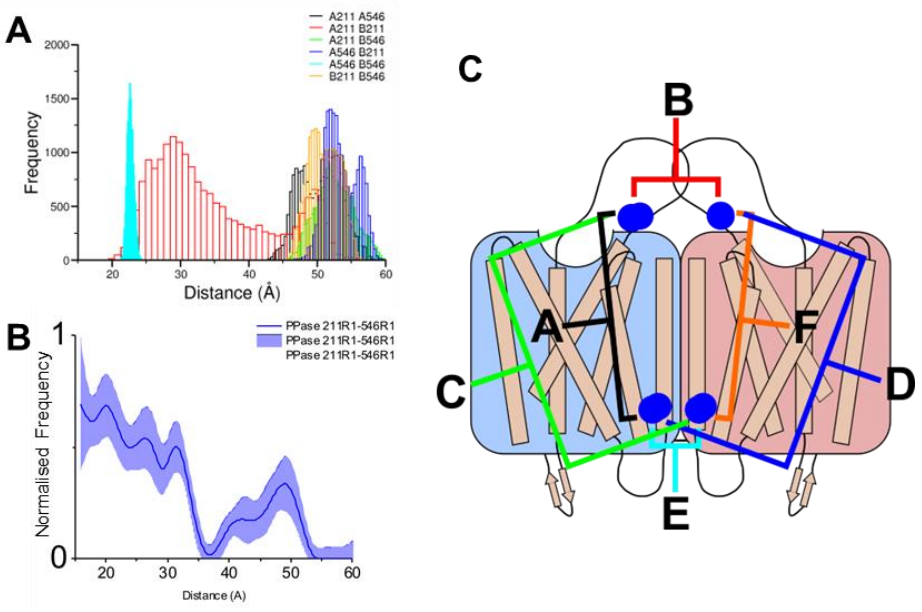


Figure 6.3 Comparison of *in vitro* and Simulation-based Dynamics
Comparison of the resting state dynamics in **A)** simulation and in **B)** PELDOR alongside **C)** the relevant distances from the simulation analyses.

the simulation demonstrated that the distance between these two residues was very variable, but there were populated states at 23 Å, 25 Å, 29 Å and 52 Å with a shoulder from 45 Å distances, compared to a similar level of flexibility with peaks at 20 Å, 27 Å, 31 Å and 49 Å with a shoulder at 42 Å as seen in the PELDOR experiments. The couple of small differences between the peaks could be explained by the simulation work measuring from the C_α of the residue rather than the side chain/spin-label. The benefit of the MD simulations here was that each peak could be attributed to different distances, for example the 23 Å peak is composed primarily of the E distance, whereas 25 Å and 29 Å are due to the B distance.

These analyses confirm what was seen in a previous study on a 100 ns timescale (Shah *et al*., 2017), that this stage of the catalytic cycle is dynamic and flexible, particularly the 5-6 loop. This is logical for this state, as there are no small molecules or Mg^{2+} cage to stabilise the protein conformation and it is open for Mg_2PP_i to enter the active site. The high flexibility of this region also provides an explanation for its absence in the crystal structure, as it was likely too dynamic to resolve. In addition to understanding the protein dynamics, attention to the motions of helices compared to the crystal state allowed identification of changes in arrangement during simulation and revealed a potentially more physiologically representative mPPase resting state. Additionally, the agreement between the dynamics and distance measurements, as seen by PELDOR and FRET, and my simulations provides validation of these approaches and their contribution to the field.

6.2.1.2 IDP- and ETI-Bound States have Limited Dynamics

The substrate-bound state was of particular interest, as it is the point at which questions around the coupling of ion pumping and hydrolysis may be answered. The small molecules in this work were IDP, the binding of which has been examined through structural studies, which provided the starting configuration for this work. In addition to IDP, another substrate analogue, etidronic acid (ETI) was also investigated. There are no available structures of ETI binding, and so the binding of IDP was used as a starting configuration for this molecule. Neither IDP nor ETI are endogenous substrates, but are non-hydrolysable analogues, and were chosen for study as the data could be compared to *in vitro* work.

The IDP and ETI symmetrically occupied states had highly similar RMSD/C_α values compared to the modelled structures, both reaching ~2 Å (**Fig. 6.2**). This was only slightly lower than what was seen for the resting state, but comparison of the RMSD/C_α pre- and post-simulation demonstrated that these structures are stable (1.3 Å) compared to the resting state (2 Å), which was further confirmed by the RMSF/C_α analysis. This demonstrated that the 5-6 and 13-14 loops were far more stable in their conformations in this state rather than in the resting state of the catalytic cycle. This change for the 5-6 loop is easily attributed to the interaction between $E^{5.76}$ and the M2 ion in the catalytic site. The overall lack of dynamics in this state was indicative of the tight packing of the occupied subunits and extensive interactions between the content of the active site and the surrounding protein (**Fig. 2.1C**). There are two possible explanations for this behaviour. Firstly, the conformation may need to be maintained for an extended period of time to allow hydrolysis to occur, particularly as mPPases are known to have ineffective nucleophilic water activation (Lin *et al*., 2012). Alternatively, this symmetrical conformation may be “locked” and not dynamic, as previous work has shown that states with both subunits occupied by ligand have a decreased rate of hydrolysis (Artukka *et al*., 2018). Despite the only small conformational changes in the substrate-bound state during simulation, again the arrangement of TMH12 changed, as the helix bends at the level of $A^{12.47}$ but maintains its positioning at either end of the helix. This indicated that the interaction between $D^{6.43}$-$K^{12.50}$ was not maintained, as in the resting-state simulation, but the lack of motion of other TMH restricts the motions that can be made upon loss of this interaction.

One of the questions these systems were designed to answer was whether differences between the ETI- and IDP-bound states could be identified. Electrometric studies of Na^+ pumping triggered by PP_i, IDP and ETI suggest that IDP is a better substrate mimic than ETI (Li *et al*., 2016; Shah *et al*., 2017). However, FRET studies indicated that ETI-bound mPPases may act more similarly to the PP_i-bound protein than IDP-bound (unpublished data obtained by Dr Christos Pliotas through PELDOR measurements). Due to these data, it was suggested that IDP may arrest the catalytic cycle at an intermediate step, rather than a rate limiting step,

and so these differences might be explained at a molecular level through these simulations. However, the IDP- and ETI-bound structures have almost identical dynamic profiles in this work. This may be due to the timescale of the simulations, which may not have been long enough for the protein to adapt to a new molecule, or an indication that the magnesium ions in the active site contribute to protein dynamics over the identity of the substrate analogue.

However, there is also the likely explanation that the ETI-bound system did not represent the *in vitro* situation. There are crystal structures for multiple IDP-bound states (Kellosalo *et al*., 2012; Li *et al*., 2016; Vidilaseris *et al*., 2019), but during this work there were no available ETI-bound structures. Therefore, the ETI was positioned through superimposing the molecule with the IDP structure, this was assumed to be reasonable as the molecules have many similarities; both have two phosphate groups linked by a central backbone atom, either nitrogen in IDP or carbon in ETI. However, the ETI backbone carbon also possesses a methyl and a hydroxyl group, whereas the central nitrogen of IDP only coordinates a hydrogen. The additional ETI groups mean that there are two possible orientations in the active site, and the positioning of the hydroxyl may be essential for properly mimicking the backbone oxygen of PP_i. Preliminary studies also suggest that ETI may prevent occlusion of the active site (Kellosalo *et al*., 2012), which is not represented in these systems.

6.2.1.3 There are State-Dependent Interactions Between the 5-6 and 13-14 Loops

The marked difference in the dynamics of the 13-14 loop between the resting state and the substrate-analogue bound state was of particular interest in this work, as TMH 14 may be involved in mPPase asymmetry and inter-subunit communication (Vidilaseris *et al*., 2019). Additionally, as TMH 13 and 14 are in the outer ring of the mPPase, they form interactions with the lipid bilayer, which the previous chapter argued could be important for functional modulation (Holmes *et al*., 2022). The change in dynamics in this region are indicated in the crystallographic models of the resting and substrate-bound states, as the loop is not resolved in the resting state, likely due to high flexibility, and there are pronounced changes in helical arrangement at the cytoplasmic end of the helices in the IDP-bound structure compared to the resting state (Li *et al*., 2016). TMH 13 bends towards the subunit interface, and TMH14 displays unwinding of the helix (**Fig. 6.4**). It was previously hypothesised that the rearrangement of the 13-14 loop served to further stabilise the 5-6 loop closure (Li *et al*., 2016). However, how this interaction is selective to the substrate-bound state has not previously been investigated.

Distance analysis between the C_α of the central residues on each of these loops ($E^{5.76}$, $D^{5.77}$ and $D^{5.78}$ of the 5-6 loop and $L^{13.71}$, $L^{13.72}$, $E^{13.73}$ of the 13-14 loop) in both the resting state and IDP-bound state systems confirmed the crystallographic conclusion that the 13-14 loop closes

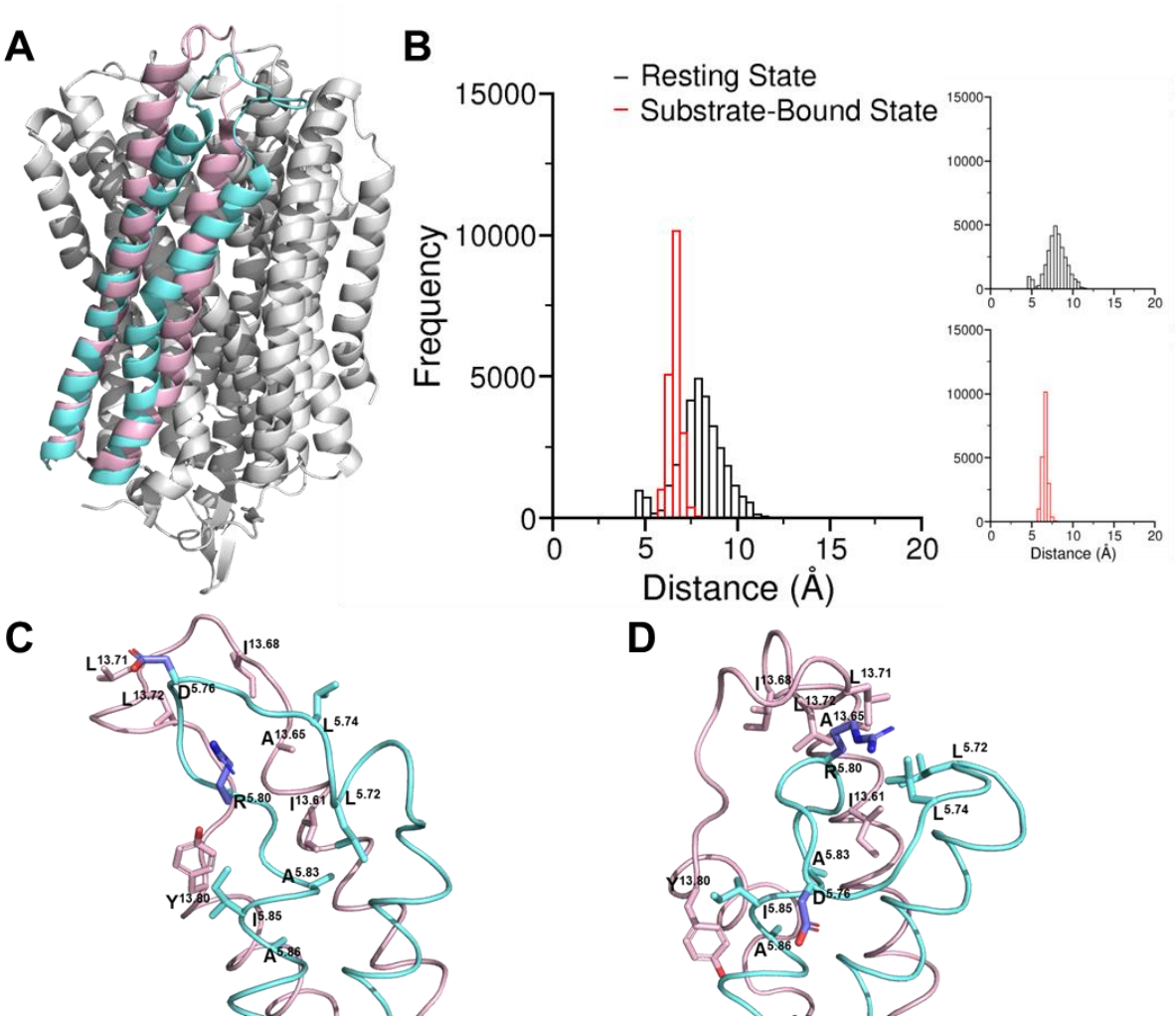


Figure 6.4 Simulations uncover state-dependent interactions between the 5-6 and 13-14 loops
A) The differing conformation of TMH 13 and 14 between states following 250 ns simulation, where light blue denotes the resting state, which bends away from the dimer interface, light pink represents the IDP-bound state, which straightens towards the interface and has a structurally resolved loop structure. **B)** Distance histograms displaying the change in proximity of the C_αs of the middle residues of the 5-6 and 13-14 loops in the resting state and IDP-bound state. **C-D)** the conformation of the 5-6 and 13-14 loops in the resting and substrate-bound state, respectively. The 5-6 loop is shown in light blue and the 13-14 loop in light pink, and the hydrophobic residues that form a solvent inaccessible interface are shown as sticks, and the charged $D^{5.76}$ and $R^{5.80}$ residues are shown in purple.

over the 5-6 loop and stabilises it (**Fig. 6.4**). Analysis revealed that these loops maintained their arrangement through the formation of a solvent inaccessible interface between the hydrophobic residues of each loop. The presence of this interaction in only the substrate-bound state is seen in the distance analysis of the loops in the resting state, as they do not maintain close contact during the simulation, other than a transient connection between the sidechain of $E^{13.73}$ and the mainchain of $D^{5.78}$. A hypothetical mechanism for the formation of this interface is as follows: while in the resting state, the 5-6 and 13-14 loops move and rotate freely, but upon binding of Mg_2PP_i, the positive charge in the active site repels the $R^{5.80}$ and attracts $E^{5.76}$, resulting in a twisting of the 5-6 loop, thereby exposing the hydrophobic residues $L^{5.72}$, $L^{5.74}$, $A^{5.83}$, $I^{5.85}$ and $A^{5.86}$ to the solvent. The hydrophobic residues $I^{13.61}$, $A^{13.65}$, $I^{13.68}$, $L^{13.71}$, $L^{13.72}$, $Y^{13.80}$ on the 13-14 loop interface with the 5-6 loop to stabilise this state. This may be

responsible for the pushing of TMH 13 and 14 toward the subunit interface, so that the inter-loop hydrophobic interface can be maintained. However, due to the lack of spontaneous formation of this in the resting state simulations, or disruption of this in the substrate-bound simulations this can only be hypothesised and requires further investigation. The formation and maintenance of this interaction may be involved in intersubunit communication, as it could affect the arrangement of TMH 13 and 14 with relation to the dimer interface.

6.2.1.4 The Product-Bound State is Highly Dynamic

There are structures of the mPPase product-bound state, bound to both of the PO_4 products and where one has left (Kellosalo *et al*., 2012; Li *et al*., 2016). The *Tm*-PPase structure contains both phosphate groups and the 5-6 loop is closed over the active site (Kellosalo *et al*., 2012). Experimentally, it is challenging to study the product bound state, as the state is hard to maintain for an extended period of time, as the PO_4 product can be converted to PP_i overtime by mPPases. Therefore, MD simulations were thought to be well placed to study this period of the catalytic cycle, particularly as no simulations have been performed of the product-bound state prior to this work.

Simulation of this state resulted in higher RMSD/C_α values than the other states investigated in this work (**Fig. 6.2**), reaching around 3.5 Å, the RMSF/C_α analysis revealed that all regions of the protein contributed to these substantial changes, but in particular the 5-6 and 13-14 loops had RMSF/C_α values approaching 10 Å. The flexibility of the 5-6 loop was seen in the resting state system, but the large movements of the 13-14 loop was not seen in this state. Upon visual inspection of this region of the protein, it was clear that during simulation, large scale changes to the structure of these helices had occurred. TMH 13 unwound by 3 helical turns and some of the helical structure of TMH14 was lost at the cytoplasmic end, this goes some way to explain the increased dynamics of the 13-14 loop as the increased length allowed more range of motion. Despite the increased dynamics, the arrangement of the 13 and 14 helices resembled the substrate and product bound state, rather than starting to transition to the resting state, as the helices remained towards the subunit interface. This was confirmed through measuring the RMSD/C_α relative to the resting state crystal structure and the post simulation resting state, which did not demonstrate structural convergence.

Due to the lower structural quality of this crystallographic model, this may explain the large dynamics during simulation, as several of the loops between helices are missing – in particular the 5-6 and 13-14 loops. Therefore, it may be that the modelling of these missing regions prior to simulation did not fully represent the native state and the structure was unable to recover during simulation time. However, if these dynamics are reliable, they may provide explanations for the reduced resolution of the product-bound states available, both of which are 4 Å

resolution. If these states are highly dynamic then they are difficult to resolve crystallographically, and the motion at the crystallographic side of the protein may impede lattice formation during crystal growth (Rosenbaum *et al*., 2007).

6.2.2 Simulations of the Resting State Reveals the Pathway of Sodium Ions to the Ionic Gate

Whilst investigating the dynamics of the resting state, sodium ions from the solvent were seen to enter the protein and travel towards the ionic gate during the 250 ns of simulation time (**Fig. 6.5**). In accordance with the molecular mechanism, the pumped sodium ion must enter prior to the Mg_2PP_i binding in Na^+-pumping mPPases (Kajander *et al*., 2013), but the resting state crystal structure does not reveal the position of the pumped ion (Kellosalo *et al*., 2012). Therefore, my simulations revealed for the first time the pathway of the ion to the ionic gate and the start of the ion translocation pathway.

It was seen that the Na^+ predominantly entered through the catalytic centre, traversed the coupling funnel and interacted with residues known to coordinate Na^+ in other catalytic states ($D^{6.50}$, $S^{6.54}$, $N/D^{16.46}$ and $K^{16.50}$), with the exception of $E^{6.53}$ (Li *et al*., 2016). Analysis of the Na^+ density during the replicas demonstrated that, with one exception, the ions followed a similar path to the ionic gate in each of the subunits in each of the replicas. In this exception, the ion did not follow this path, rather entering between helices 8 and 9 and subsequently not interacting with the Na^+-binding residues. Therefore, as it was unlikely to accurately represent the actual mode of Na^+ entrance to the protein and transition to the binding site, it was excluded from the following analyses.

Contacts analysis between the protein and the binding sodium ion and 3D ion density analysis was used to understand at molecular detail the pathway of the ion to the ionic gate. Initial inspection of this data indicated that there were transitional binding sites in the active site, as the sodium ion occupied the region above $R^{5.50}$ for an extended period of time. However, the lack of significant interactions with any residues in this area indicated that the ion was not being coordinated in the active site, as the residue identified through contact analyses was $R^{5.50}$. Further analysis indicated that rather than being transiently bound, the passage of the sodium ion to the ionic gate was being blocked by the arginine, causing its occupancy of the active site. These simulations were used to identify the mechanism by which sodium passes the positively charged arginine (**Fig. 6.5**). The sodium is first coordinated by $D^{16.46}$, this stabilises the location of the ion in the presence of $R^{5.50}$ and can result in recruitment of $D^{6.50}$ for coordination by both aspartates. Once this double coordination is in place, the anionic side chains reorientate and move the sodium ion towards the ionic gate and the semi-conserved glutamate, while forcing $K^{16.50}$ to rearrange and accommodate the cation.

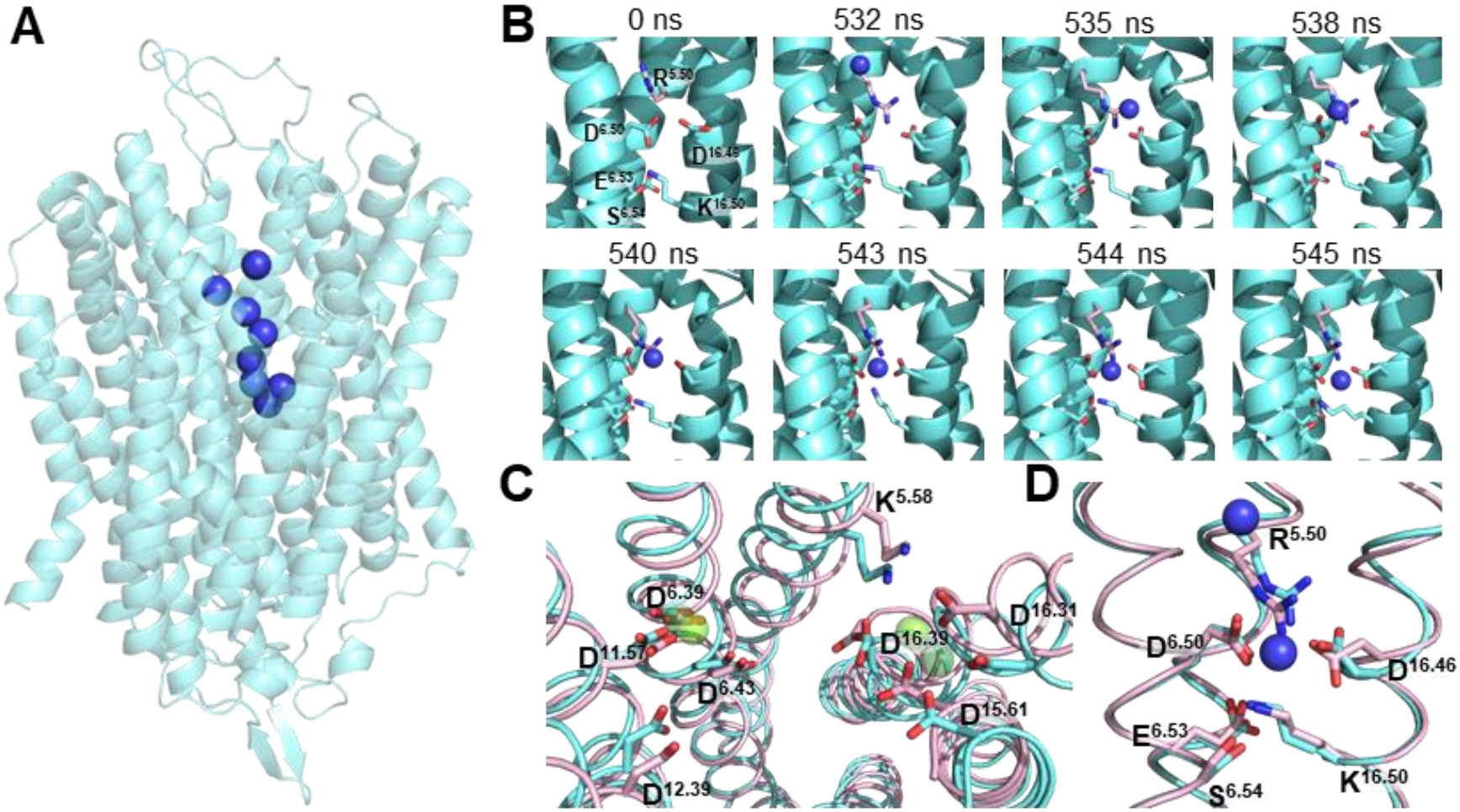


Figure 6.5 The sodium binding mechanism and subsequent conformational changes were revealed through simulation
The **A)** The pathway of the sodium ion (blue sphere) through the active site and coupling funnel to the ionic bridge shown through representative frames. **B)** The transition of the ion past residue $R^{5.50}$ to the ionic site through interactions with $D^{6.50}$ and $D^{16.46}$, and displacement of $K^{16.50}$ by the ion. **C)** The conformational changes at the active site caused by ion binding to the ionic gate. Light blue represents the starting structure and light pink the structure at 544 ns. For clarity the active site Mg^{2+} ions are transparent spheres, dark green for the initial structure and light green at 544 ns. **D)** The conformational changes at the ionic gate following ion binding.

The localisation of the sodium ions at the end of the simulation was similar between replicates, but in most cases somewhat different to that seen in the IDP-bound crystal structure (Li *et al*., 2016). They were displaced by an average of 3.16 Å from the IDP-bound site, typically in the Z direction, which may explain the lack of interactions with the semi-conserved glutamate, since it is lower down helix 6. There was no correlation between the length of time in the binding site and the positioning of the ion compared to the IDP-bound structure. This suggests that conformational changes following the binding of Mg_2PP_i may cause the sodium to shift downwards to the ionic gate. In particular the twisting and downward motion of TMH 16 as seen in the substrate-bound crystal structure (Li *et al*., 2016), may bring the sodium closer to the semi-conserved glutamate and into the crystallographically defined binding site. Although, it should be noted that the positioning of the sodium ion in the simulations of the substrate-bound state reported here changes throughout the simulation time. Typically, a movement upwards in the Z direction to more closely resemble the positioning seen in the sodium-bound resting state.

Binding of the pumped ion itself caused conformational changes in the protein, both at the ionic gate and in the active site (**Fig. 6.5**). At the ionic gate, TMH 16 contracts towards TMH6 and the $N/D^{16.46}$ residue sidechain reorientates to better coordinate the cation, resulting in the ionic gate resembling the substrate-bound arrangement. These movements at the ionic gate resulted in rearrangements in the active site. Of the inner ring of helices, TMH 11 and 12 did not exhibit changes in arrangement, but the TMH 5 and 6 exhibited small movements upon sodium binding. TMH 6 moved further into the active site and TMH 5 twisted by 1.9 Å (as measured from the C_α of the $K^{5.58}$ residue at time 532 ns and 544 ns). Likely due to this twisting of TMH 5, the neighbouring TMH 16 leans towards 5 by 2.6 Å (as measured from the C_α of the $T^{16.40}$ residue at time 532 ns and 544 ns) and increasing to 3 Å towards the cytoplasmic end of the helix (as measured from the C_α of the $D^{16.31}$ residue at time 532 ns and 544 ns). Despite this, the $D^{16.39}$ residue positioning is maintained through a slight loss of helix geometry on this helical turn. TMH 15 also moves by 2.5 Å in this direction (as measured from the C_α of the $D^{15.61}$ residue at time 532 ns and 544 ns), following TMH 16. Overall, this had the effect of making the active site more compact and altering the position of some of the key active site residues ($K^{5.58}$, $D^{15.61}$ and $D^{16.31}$).

This was an interesting finding, as it has been hypothesised that conformational changes upon binding of the sodium may be crucial for promoting PP_i binding, as serial crystallography trials saw that, in the absence of Na^+, PP_i density could not be identified at short soaking times (Strauss, 2021). Therefore, this rearrangement of the active site could be beneficial for promoting Mg_2PP_i binding.

6.2.2.1 An Updated Understanding of Sodium Binding to mPPases

Understanding the pathway of the sodium ion to its binding site and how it binds could be key to understanding the cation pumping specificity of different mPPase subfamilies. Na^+-pumping mPPases have a semi-conserved glutamate at position $E^{6.53}$ to form an ion triple with $D^{16.46}$ and $K^{16.50}$ (Li *et al*., 2016; Vidilaseris *et al*., 2019), whereas H^+-pumps only have the DK ion pair and the glutamate is displaced one helical turn to $E^{6.57}$ (Lin *et al*., 2012). Mutagenesis to change the position of the glutamate has been seen to convert an Na^+-PPase to a H^+-PPase (Luoto *et al*., 2011). Therefore, the semi-conserved glutamate was thought to be key to understanding cation specificity in mPPases (Tsai *et al*., 2014). However, the reverse mutation does not yield a Na^+-PPase from a H^+-PPase, nor account for dual Na^+/H^+-pumps, which also possess the ion triple configuration at the ionic gate (Li *et al*., 2016).

The findings of these simulations provide an updated understanding of ion binding and conformation of the ionic site prior to substrate binding, compared to that outlined in 2016 (Li *et al*., 2016). In this structural interpretation, it was proposed that the movement of TMH 12 downwards and the loss of the $D^{6.43}$-$K^{12.50}$ ion pair were what caused the correct positioning of the ionic gate residues, ultimately concluding that the Na^+ binding site only forms upon substrate binding. This work alters some previous assumptions in the mPPase field, as loss of the $D^{6.43}$-$K^{12.50}$ ion pair may occur prior to sodium binding, allowing movement of the cytoplasmic end of TMH 12, but not the downward motion described structurally. Nevertheless, the ionic gate residues do form the ion binding configuration seen in the substrate-bound state, but there is not full coordination of the ion as described structurally. Therefore, it is likely that substrate binding triggers further ionic gate rearrangements, and so these simulations were unable to provide further clarity to the role of the semi-conserved glutamate in cation selectivity.

The mechanism of sodium binding as predicted by atomistic simulation is as follows: the sodium approaches the entrance to the mPPase active site and diffuses into the active site which it occupies for some time. However, the ion is not bound by specific residues, rather it is unable to continue its path to the ionic gate due to electrostatic repulsion by $R^{5.50}$. This is overcome by coordination by both $D^{6.50}$ and $D^{16.46}$, which change from an upward facing arrangement, to a downward one to bring the cation to the ionic gate. This motion causes rearrangement of TMH 16, which moves closer to TMH 6, triggering a contraction of the cytoplasmic end of the helix. This causes the active site to change through further contraction of TMH 15 and rotation of TMH 5, which may promote Mg_2PP_i binding. Upon this binding, the ionic gate may further contract as seen in the substrate-bound structure, thereby causing further downward motion of the Na^+ towards the semi-conserved glutamate.

6.2.3 Investigation of Conformational Changes Between States

The previous systems discussed in this chapter involved the simulation of structurally defined states but understanding the transitions that occur in the protein between states was still of interest. Therefore, systems were set up to investigate whether MD simulations could be used to understand and capture conformational changes between states of the mPPase catalytic cycle, through addition or removal of active site contents to trigger changes. This would be particularly beneficial to the field, as, despite many of the catalytic states having been structurally defined, important transition states are missing from our understanding. Here, I attempted to trigger the change between different conformational states through manipulation of the active site contents or application of steering forces.

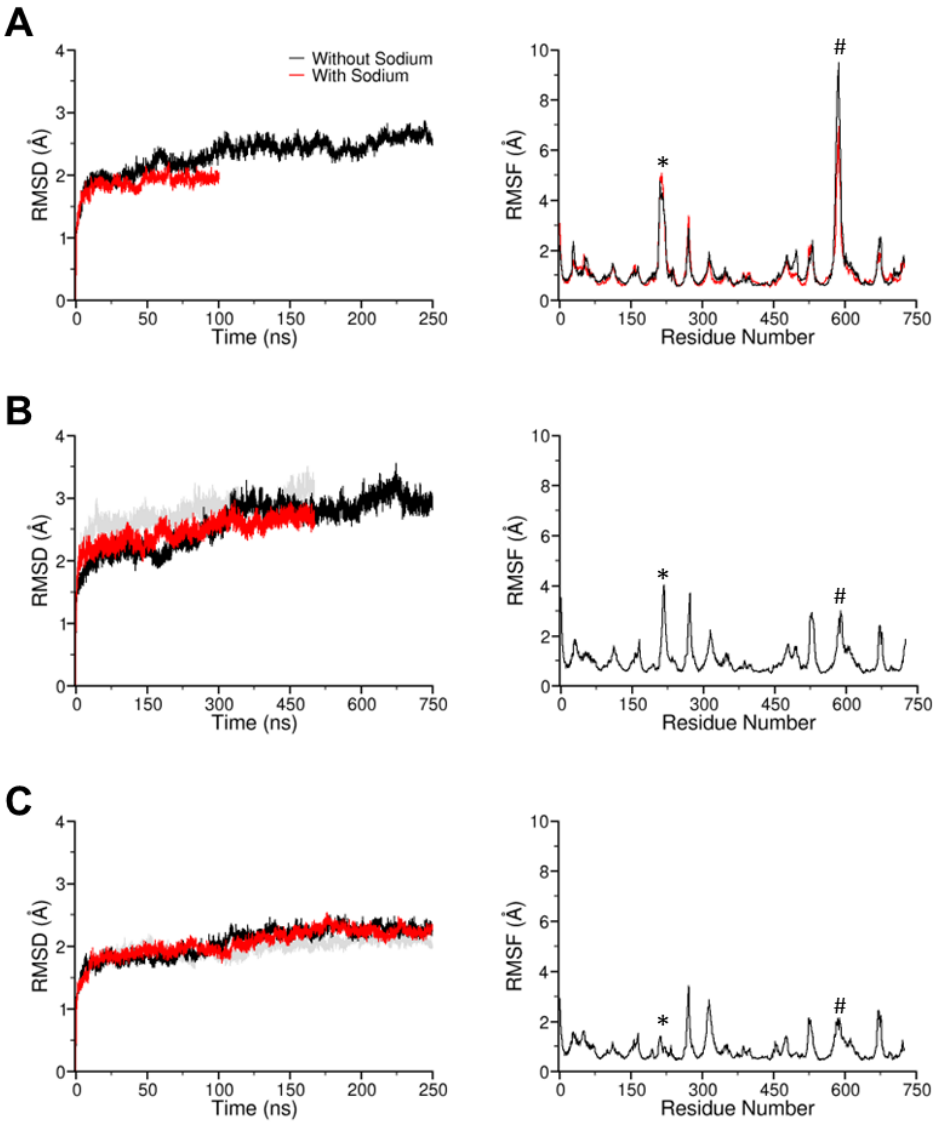


Figure 6.6 The dynamics of hypothetical states and those used to study transition between states
The RMSD/C_α of the helices and the RMSF/C_α of the whole protein averaged across replicas and subunits of **A)** the resting state in presence of IDP and Mg^{2+} ions to trigger transition to the substrate-bound state. Without and with sodium denotes the presence of sodium at the ionic gate at the start of the simulation. **B)** The "empty" substrate-bound state, where all small molecules and ions were removed with the exception of the Mg^{2+} ions found in the resting state, to trigger relaxation of the substrate-bound state. **C)** The asymmetric IDP-bound state, where one subunit resembled the "empty" substrate-bound state, and the other the substrate-bound state. The * and # indicate the 5-6 and 13-14 loops, respectively.

6.2.3.1 Substrate Binding to the Resting State

Addition of IDP and the active site magnesium ions to the resting state did not prompt transition to the substrate-bound state during 250 ns. The protein dynamics remained similar to that seen in the resting state (**Fig. 6.6**), and the IDP and magnesium ions were not held tightly in the active site. This was performed in both the absence and presence of the bound sodium ion, to see whether the sodium-related structural rearrangements could be key for promoting IDP binding and transition to the next conformational state. However, there were no identifiable differences between the sodium-bound or unbound systems with relation to substrate-related conformational changes. This indicates that mechanisms for IDP binding and resulting conformational changes may be more complex than solely changes propagated from the active site throughout the protein, perhaps intermediate binding poses or interactions with the 5-6 loop.

6.2.3.2 Relaxation of the Substrate-Bound State

To assess whether the substrate-bound state could be "relaxed" into the resting state, the active site contents were removed with the exception of the two magnesium ions found in the resting state and simulated for 250 ns. This system did have larger RMSD/C_α values than the fully occupied substrate-bound system, reaching ~3 Å (**Fig. 6.6**). However, RMSF/C_α analysis revealed that this was mainly localised to the 5-6 or 13-14 loops. Despite the relaxation of the 5-6 and 13-14 loops, likely due to loss of positive charge in the active site allowing the 5-6 loop to no longer be restrained, these loops did not substantially change their positioning. Perhaps due to this, the structure did not relax to resemble the resting state during simulation time, even upon extension to 750 ns in the case of one replica. Despite this, sodium ions entered the protein and travelled to the ionic binding site, as in the resting state simulations.

6.2.3.3 Investigation of Ion Pumping from the Substrate-Bound State

Triggering ion pumping from the substrate-bound state would have been a very exciting revelation from these simulations, as it would demonstrate the pumping before hydrolysis mechanism, rather than the hydrolysis before pumping mechanism, a series of events that have been very heavily debated throughout the mechanistic investigation of mPPases. From the structural studies of mPPases, movement of TMH 12 by 2 Å downwards was identified as a key helical rearrangement and so was targeted through steered MD simulations to trigger transition to the structurally undetermined transition state. TMH12 was pulled in the direction indicated structurally. However, the significant interactions between TMH12 and the other helices on the cytoplasmic side of the protein rendered this avenue of investigation inaccessible during these simulations. This indicates that conformational changes at the cytoplasmic end of the protein, perhaps change in the active site, causes loss of interactions between TMH12 and its adjacent helices, allowing ease of downward motion.

Similarly, steering forces on the sodium ion were unable to induce ion pumping through the ionic gate and structural transition. It is hard to draw a single conclusion for why this might be, as there are a number of credible explanations. Mechanistically, there is much debate over the order of hydrolysis and ion pumping, and so perhaps these simulations were attempting to bring about ion pumping too early in the catalytic cycle. Alternatively, the symmetrical IDP-bound state I was working with may be a "locked" conformation, as binding of substrate to both subunits has been seen to be detrimental to mPPase activity. This would therefore prevent further investigation of the rest of the catalytic cycle from this starting point, as it may be a catalytic dead end, and so efforts should be made to identify and characterise more physiological catalytic structures.

6.2.3.4 Investigation of Asymmetric Catalytic States

As an alternative to the symmetrically occupied substrate-bound systems for further investigation of the catalytic cycle, asymmetry was investigated through these simulations. There is an increasing body of structural and functional evidence that the mPPase catalytic cycle contains elements of asymmetry (Artukka *et al*., 2018; Vidilaseris *et al*., 2019), which are beneficial for catalytic activity. These asymmetric states may be hard to study through traditional structural biology techniques, as the crystal structures of the substrate-bound states were determined in excess of IDP (Kellosalo *et al*., 2012; Li *et al*., 2016), as this can stabilise membrane proteins and improve crystal formation, or asymmetric structures have been solved in the presence of inhibitor (Vidilaseris *et al*., 2019). To investigate these asymmetrical properties of mPPases, systems were constructed based on hypothesised asymmetric states, in which only one of the two active sites contained a substrate analogue. Due to the lack of success "binding" IDP to the resting state, but some changes to the substrate-bound state dynamics following removal of the active site contents, the substrate-bound state was selected as the starting point of these asymmetrical studies.

However, changes in dynamics compared to the substrate-bound starting state or between the subunits was not seen during these simulations, including motion of the 5-6 and 13-14 loops (**Fig. 6.6**). This was disappointing, as I was unable to generate an asymmetric state, but does say something about asymmetry in catalysis and intersubunit communication. As when there is substrate and magnesium in one of the active sites, the adjacent subunit is unable to relax, but the mechanism of this was unable to be studied through these simulations due to lack of dynamic change.

6.2.4 High Throughput Inhibitor Screening Against Updated Structural Models

Despite mPPases representing validated drug targets for protozoan parasites and bacterial pathogens, as of yet, structure-based drug design methodologies have not been able to

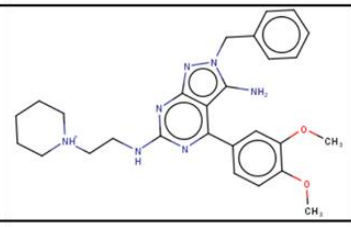
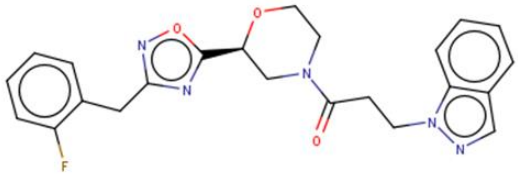
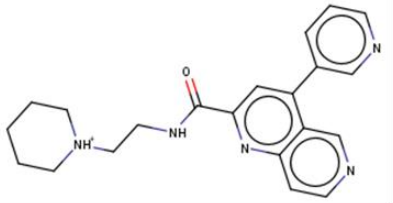
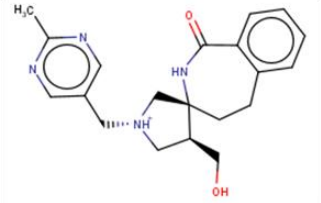
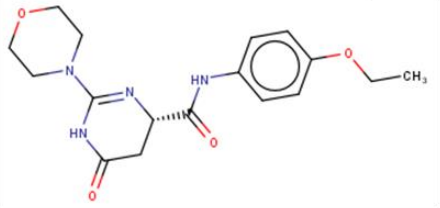
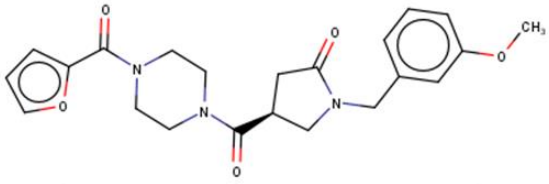
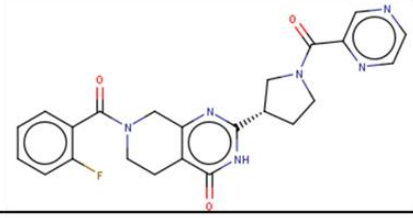
Protein Model	Compound Name	Chemical Structure	Docking Score	Interactions
Post Simulation Resting State	ALB-H09685042		-10	H-bonds: 4 Pi-stacking: 4 Salt bridges: 2 Metal: 0
	LEG-19740413		-9	H-bonds: 0 Pi-stacking: 1 Salt bridges: 0 Metal: 0
	LMK-21279597		-8.5	H-bonds: 2 Pi-stacking: 2 Salt bridges: 0 Metal: 2
	AOP-19076614		-8.4	H-bonds: 3 Pi-stacking: 1 Salt bridges: 0 Metal: 2
	9080290		-7.9	H-bonds: 2 Pi-stacking: 0 Salt bridges: 0 Metal: 2
	ALB-H11731168		-7.8	H-bonds: 0 Pi-stacking: 1 Salt bridges: 0 Metal: 2
	ASF-18956087		-7.4	H-bonds: 3 Pi-stacking: 2 Salt bridges: 0 Metal: 1

Table 6.1 Post simulation resting state identified compounds

The chemical structure and MCCB name of the top 7 molecules identified through *in silico* screening against the post 250 ns of simulation resting state structure.

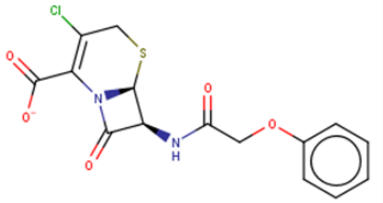
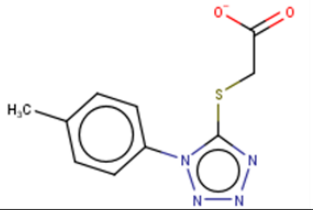
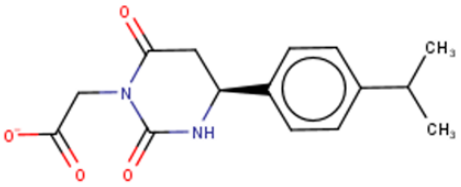
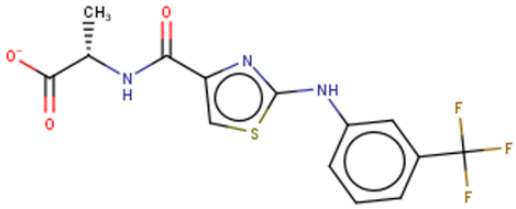
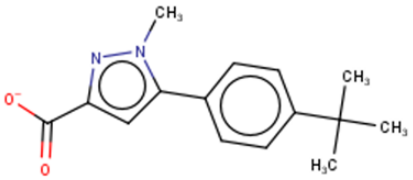
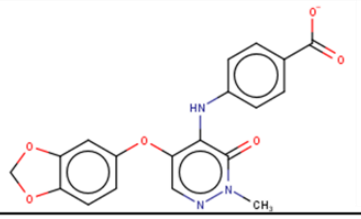
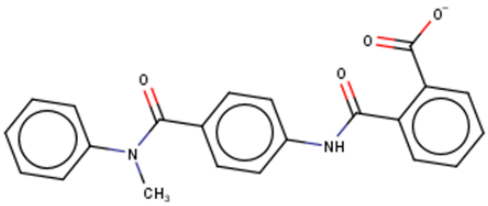
Protein Model	Compound Name	Chemical Structure	Docking Score	Interactions
"Empty" Substrate Bound State	7795842		-11.6	H-bonds: 3 Pi-stacking: 1 Salt bridges: 2 Metal: 2
	9061379		-11.5	H-bonds: 1 Pi-stacking: 1 Salt bridges: 2 Metal: 0
	ALB-H04810563		-10.8	H-bonds: 3 Pi-stacking: 0 Salt bridges: 2 Metal: 2
	ALB-H01377801		-10.6	H-bonds: 2 Pi-stacking: 0 Salt bridges: 1 Metal: 2
	ALB-H10943923		-10.6	H-bonds: 1 Pi-stacking: 1 Salt bridges: 2 Metal: 2
	ALB-H01822663		-10.5	H-bonds: 2 Pi-stacking: 0 Salt bridges: 2 Metal: 2
	9142945		-10.4	H-bonds: 4 Pi-stacking: 1 Salt bridges: 0 Metal: 0

Table 6.2 "Empty" substrate-bound state identified compounds

The chemical structure and MCCB name of the top 7 molecules identified through *in silico* screening against the "empty" substrate-bound state structure.

identify novel inhibitors of mPPase. This may be due to the methodologies used, such as unpredictive docking programmes, small available compound libraries, or due to using structures for structure-based drug design that do not fully represent the physiological behaviour of mPPases. To address the former of these problems, large external compound libraries and the in-house MCCB library, alongside a suite of docking programmes were used in this work. For the latter, novel structures from these simulations were used for computational docking to identify potential alternative binding sites and highlight their role in furthering the study of mPPases.

6.2.4.1 High Throughput in silico Docking Identified Novel Potential Inhibitors

The resting state post 250 ns atomistic simulation (descripted in section 6.2.1.1), and the substrate bound state with the small molecules removed and loop modelled open (here on referred to as the "empty" substrate-bound state") were prepared using Maestro for high throughput screening. Using both Gold and Glide programmes, the internal MCCB library of 91,000 compounds were virtually screened against the structures. Following rounds of increasingly precise docking, up to 10 binding poses for the top 20-30 selected compounds

were used to identify the compounds with high docking scores, conformational consensus across their docking poses and favourable interactions, particularly with the metal ions. This led to the selection of 7 compounds for each state (**Tables 6.1** and **6.2**), with docking scores between -10 and -7.4 for the post-simulation resting state (average -8.43), and between -11.6 and -10.4 for the "empty" substrate-bound state (average -10.86). Interestingly, despite only having an RMSD of 1.2 Å between the structures, the top 100 compounds for each structure were unique to that system, indicating that using these two states in tandem may uncover more chemical space to explore for inhibitor discovery.

Of the compounds selected, 7795842 was of particular interest, as it had the highest docking score, was seen to form the most "ugly" (denoting optimisable) interactions with the mPPase active site. 7795842 is a commercially available compound due to being a cephalosporin, a class of antibiotics that work by disrupting the formation of peptidoglycan for bacterial cell walls, through covalent binding to penicillin-binding proteins (Tipper and Strominger, 1965). Covalent bonding in this system was ruled out through computational methods. Compounds with this chemical structure lend themselves to optimisation, as they can be divided into two "halves" across the amide bond. Each of these "halves" can be optimised individually and then joined through amide coupling to assess their function, here I term them the "cephalosporin" group and the "acid" group (**Fig. 6.8**). To start optimising the acid group, fragment docking of an acid fragment library was performed, but the binding poses did not resemble those seen in the screens leading to identification of 7795842. Therefore, ROCS (rapid overlay of chemical

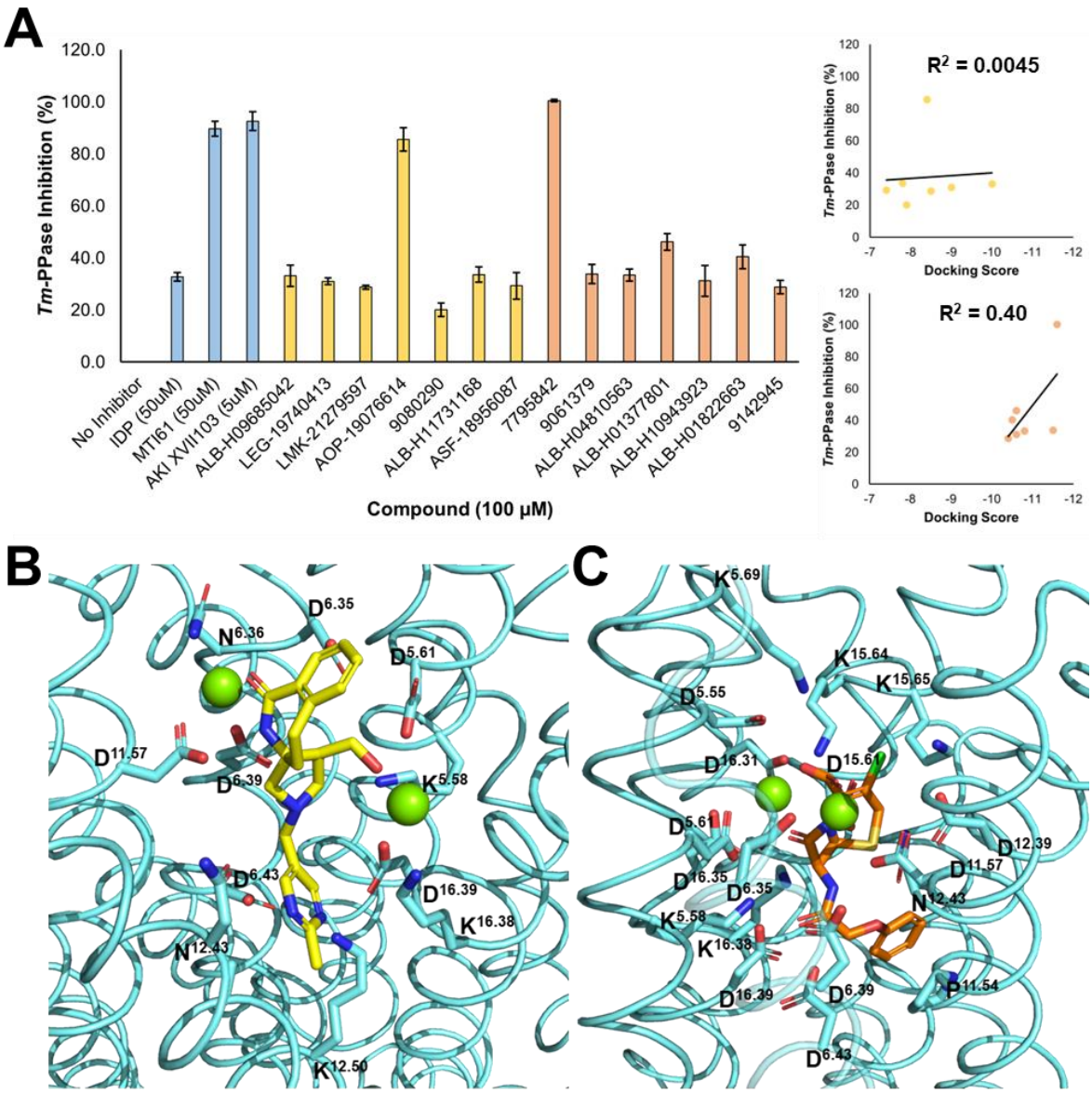


Figure 6.7 Both structures identified *in vitro* confirmed hits from *in silico* screening
A) The performance of the identified compounds *in vitro* at 100 µM, where light blue represents the controls, yellow the post-simulation resting state identified compounds and orange the "empty" substrate-bound state identified compounds. The correlation of the compound activity and docking score for each set is also shown. **B-C)** The binding poses for AOP-19076614 and 7795842, respectively, the best performing compounds. The *in vitro* data shows the mean reading (n = 3) and error bars represent the standard deviation.

structures) was used to screen the Sigma Library compounds for structural similarity with the acid group. The top 250 identified compounds were shortlisted based on their relevance and commercial availability, resulting in 73 acids. These were then coupled to the cephalosporin group and docked into the empty substrate-state, but the docking scores did not to improve compared to the 7795842 compound.

Investigation of different cephalosporin groups was performed by docking commercially available cephalosporin and structurally related penicillin compounds to the "empty" substrate-bound structure. Interestingly, this did identify compounds with higher docking scores, namely

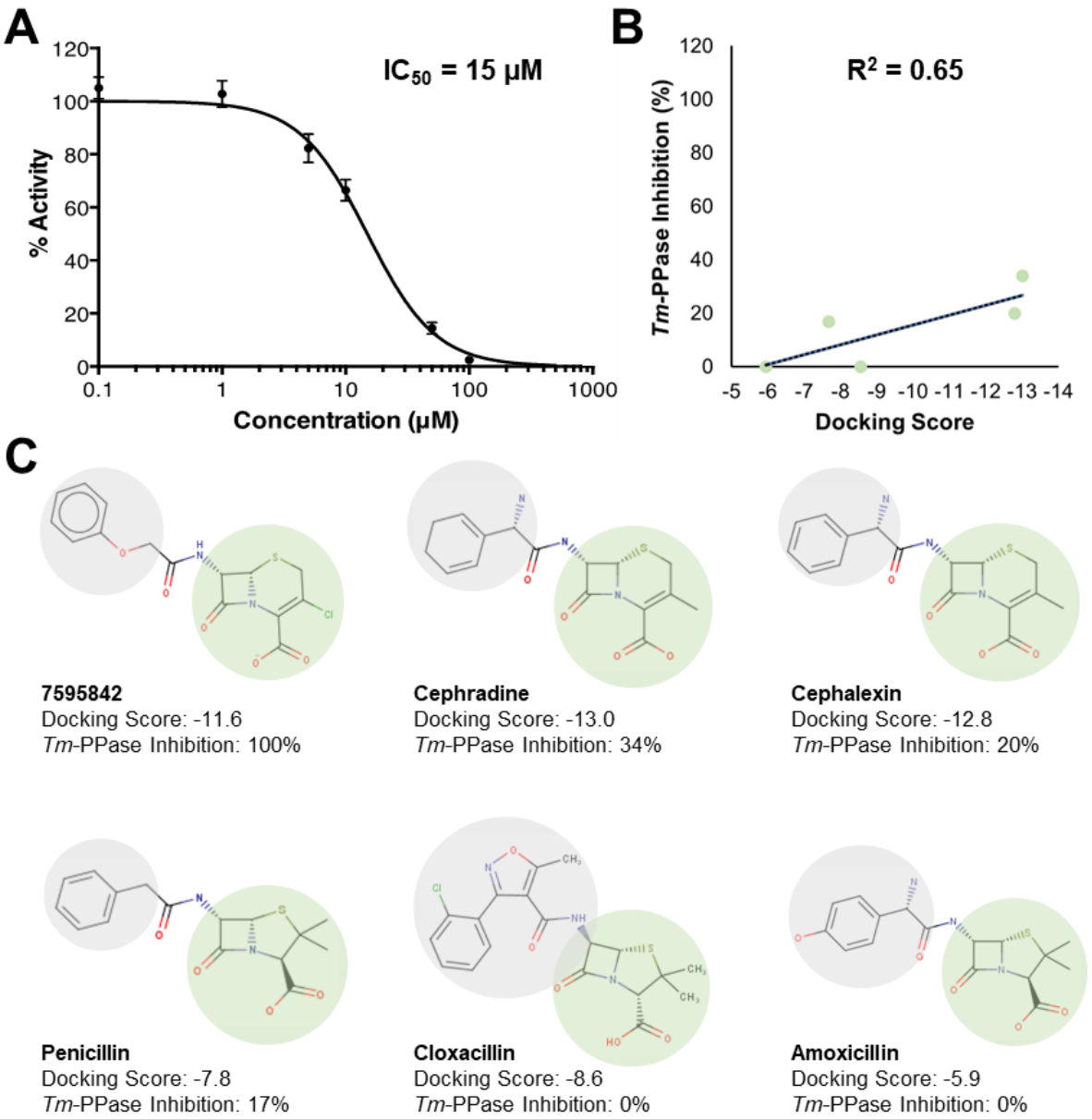


Figure 6.8 Continued investigation of the 7795842 compound and performance of similar chemicals
The **A)** A concentration-response curve of 7795842 and calculated IC_{50} for this compound. **B)** The correlation between the *in vitro* performance of related compounds and their docking scores. **C)** The structure of the 6 compounds in this series, where the different groups comprising their structure are show in light green for the "cephalosporin" group and light grey for the "acid" group.

cefradine and cephalexin, this appeared to be due to enhanced space-filling in the active site and advantageous interactions of the acid group.

6.2.4.2 Performance of the Identified Inhibitors in vitro

14 novel inhibitors identified from the post simulation resting state and empty substrate-bound state were selected for *in vitro* testing of their capacity to inhibit *Tm*-PPase activity. All of the identified compounds were able to inhibit *Tm*-PPase between 20% and 100% at 100 µM concentration (**Fig. 6.7**). In particular the compounds AOP-19076614 and 7795842 had over 80% inhibition, and so were treated as potential hits from this screening.

Compound 7795842 was carried forward into more in-depth *in vitro* investigation by Dr Keni Vidilaseris, as it possessed the highest activity at 100 µM. A full IC_{50} curve was performed between 0.1 µM and 100 µM 7795842 concentrations, which revealed that the compound has and IC_{50} of 15 µM (**Fig. 6.8**). Although this is not a low enough value to be a lead compound in the drug discovery pipeline, this is the best IC_{50} reported for novel mPPase inhibitors to date (Wilkinson, 2019), and so prompted further investigation of this potential series of cephalosporin and penicillin compounds.

Of these, amoxicillin, cloxacillin, cefradine, cephalexin and penicillin were screened *in vitro* to identify their IC_{50} values. Unexpectedly, both amoxicillin and cloxacillin had no inhibition activity against *Tm*-PPase and the IC_{50} values for cefradine, cephalexin and penicillin were 202.8 µM, 758.4 µM and 964.2 µM, respectively. Despite not completely conforming to the docking scores calculated by Glide, the activity of these compounds did correlate to their docking scores (R^2 = 0.65) (**Fig. 6.8**). This is especially useful in early structure-based compound screening, as assessing whether the *in silico* techniques can predict the *in vitro* activity is key for further *in silico* optimisation, as outlined above. These data did indicate that *Tm*-PPase preferred the 6-member ring structure of 7795842, cefradine and cephalexin over the 5-member ring structure of penicillin, cloxacillin and amoxicillin. Additionally, there appears to be a preference for flat rings in the cephalosporin group, which is useful for further inhibitor design.

6.2.4.3 MD Simulations of the 7795842-Docked State

As there was limited predictive accuracy between the *in silico* docking and *in vitro* activity assays, it was hypothesised that 7795842 binding to *Tm*-PPase captured an intermediate state in the catalytic cycle that was not fully represented by the “empty” substrate-bound structure. This was investigated through MD simulations of the docked structure with the goal of identifying conformational changes in the protein and generating an optimised structure for virtual screening. Following 100 ns of simulation time, it was concluded that simulation may not be able to identify the 7795842-bound conformation of *Tm*-PPase, as the compound exited the active site in the Z direction and did not interact with the Mg^{2+} ions as predicted *in silico*. Therefore, further investigation of the 7795842 compound series was halted until further structures became available.

6.2.4.4 Simulated Structures Performed Better than Crystallographic Structures for Predicting Inhibitory Compounds

I chose to use these novel states that have not been structurally defined, as previous *in silico* chemistry has only been done targeting the crystal structures, predominantly of the resting state (Wilkinson, 2019) and to demonstrate the usefulness of simulation generated states in

further mPPase understanding. As outlined previously in this chapter, the crystal structures available for structure-based drug design may not be fully representative of the *in vivo* state of mPPases, or not catalytically relevant. This may particularly affect the resting state, as upon sodium binding, which is likely the physiological status of resting mPPases, there are structural rearrangements in the active site that impact virtual screening. Additionally, small molecule binding could induce conformational change to a substrate-bound like state, and so using this for screening seemed like a way of optimising the screening towards active compounds and investigate more of the protein landscape for inhibitor identification and design.

The post simulation resting state and empty substrate-bound states both performed well in virtual screening, as both produced *in vitro* verified hits. However, the empty substrate-bound structure appeared to perform better than the post simulation resting state, as it identified compounds with better docking scores (-10.86) compared to the resting state (-8.43), and in the *in vitro* assays, the compounds generated against the empty substrate-bound model appeared to have a slightly higher average inhibition (44.9%) compared to the resting state crystal structure (37.4%). Due to this, the docking scores were compared to *in vitro* activity, which further confirmed that the “empty” substrate bound state appeared to perform better for structure-based compound screening, as there was a higher correlation between docking score and activity (R^2 = 0.402), compared to the resting state (R^2 = 0.0045). Although this correlation is still objectively low, in the context of docking tools, any level of increased predictability should be exploited.

These data indicate that the resting state may be a poor starting point for structure-based drug design. It is highly flexible, as shown in these simulations, and so may be difficult to stabilise through the activity of small molecules. Additionally, it is likely that binding to this state may cause a conformational change and so the structure does not provide a suitable platform to use for further optimisation of the chemicals. However, the modelled “empty” substrate-bound state performed very well, perhaps because the contracted inner helices and the closeness of the two Mg^{2+} ions might better resemble an mPPase post-inhibitor binding and allow predictive poses or force the docking programmes to identify compounds that will induce mPPases to occupy this state. This may explain why the “empty” substrate bound state performed well, but the inability to generate further compounds in the 7795842 family or improve its activity indicates that this molecule may not be binding exactly as predicted by the *in silico* methods and instead forming another state.

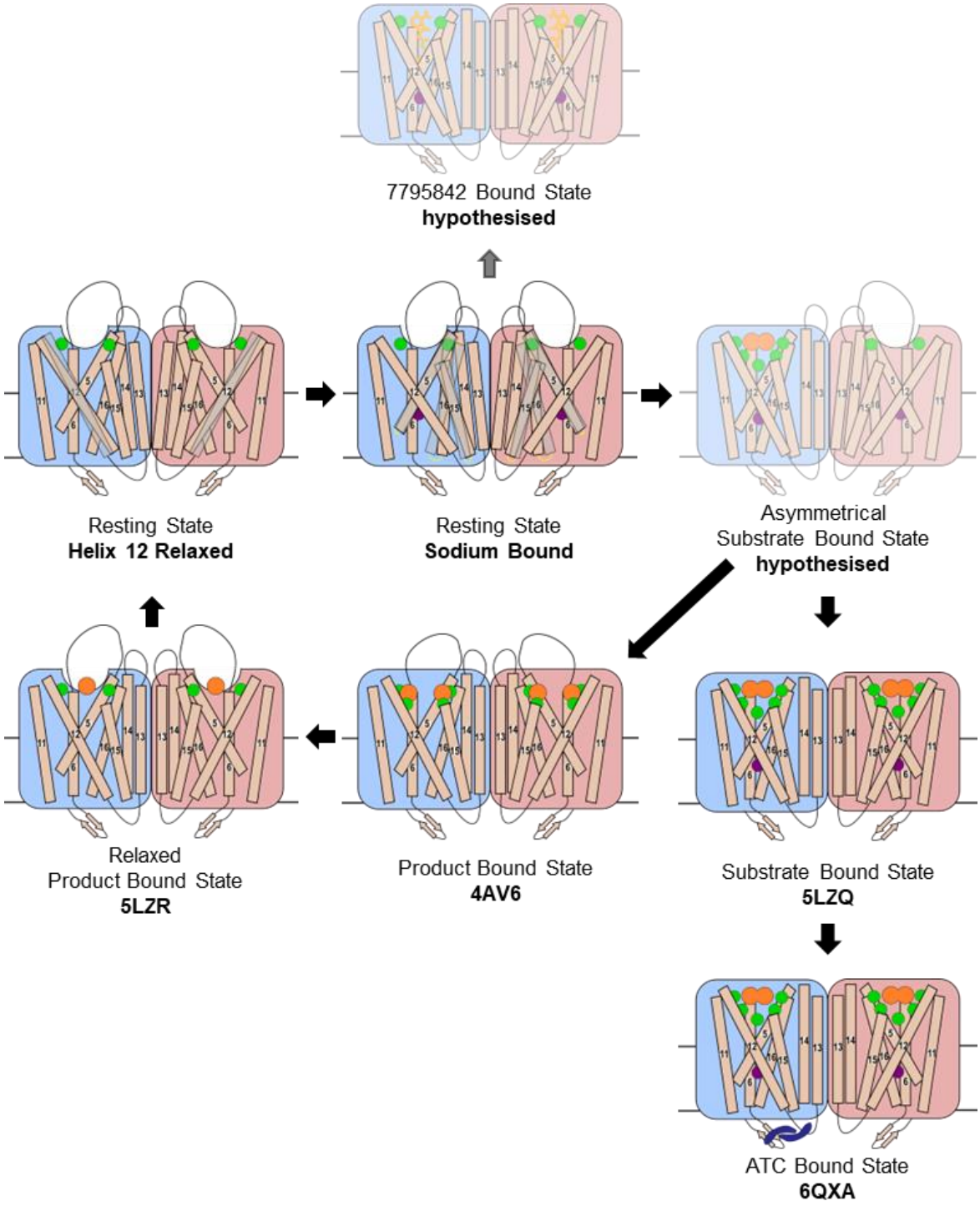


Figure 7.9 An updated understanding of the *Tm*-PPase catalytic cycle
The work presented in this chapter provides an updated understanding of the catalytic cycle and structural details of these states. The resting state adopts a different conformation in simulation to the published structure and so alters understanding of the TMH 12 motions. Transitions between states and the asymmetry in catalysis are still poorly defined, but this work adds to the evidence that the crystallographic substrate-bound structure is perhaps inactive, and the asymmetrical structure progresses to the product-bound states, which are highly dynamic. In addition to these, another inactive state, bound to the small molecule 7795842 has been hypothesised in this work, and may provide additional understanding of the early stages of the cycle. Mg^{2+}: green spheres, Na^{+}: purple spheres, PP_i and PO_4: orange spheres, ATC: blue dimer, 7795842: orange sticks. Updated helical arrangements are shown, with previous arrangements in grey.

6.3 Conclusion

The work presented in this chapter has provided significant contributions to understanding the catalytic cycle of mPPases. I have, for the first time, uncovered the pathways of the pumped sodium ion to the ionic gate, including the mechanism by which it passes $R^{5.50}$ through coordination by $D^{6.50}$ and $D^{16.46}$, and the resulting conformational changes upon binding. These changes included twisting at the ionic gate and movement of helices and residues that directly coordinate substrate and magnesium ions in the active site. This provides insight into the early steps of the catalytic cycle (**Fig. 6.9**), and a more physiologically representative view of resting state mPPase structures, as they are likely predominantly sodium bound. Additionally, simulation of the resting and substrate-bound states demonstrated the difference in dynamics between them, but also motion of TMH 12 in both states, suggesting loss of an ion pair compared to structural studies. This also allowed the identification of state-dependent interactions between the 5-6 and 13-14 loops that seem to "lock" the substrate-bound conformation. This interaction is maintained through formation of a hydrophobic interface between the two regions. Using these insights and simulation generated structures, computational chemistry was carried out and confirmed the ability of these structures to perform better in *in silico* screening studies than their crystallographic counterparts, and therefore their capacity to contribute to the continued study of this family of proteins. This also provided some experimental confirmation of the reliability of these states, alongside PELDOR measurements of mPPase dynamics. However, various parts of the catalytic cycle were inaccessible for study through MD simulations, for example structural transitions and asymmetry, but sharing this with the field is essential for efficient continued efforts to understand these clinically relevant proteins.

7. The Role of the Blade in mPIEZO1 Function

7.1 Introduction and Rationale

mPIEZO1 is a large complex protein, made up of three identical subunits, each composed of a long blade of 36 transmembrane helices arranged into four-helix transmembrane helical bundle units (HBU), some of which have intracellular membrane parallel helices (MPH) projecting from them in the XY plane. Additionally, under HBU7, 8 and 9 sits a long (16 helical turns) beam helix that follows the direction of the blade (Guo and MacKinnon, 2017). The subunits interact primarily *via* the central pore, which is made up of two domain switched transmembrane helices (TMH) from each subunit, resulting in three inner and three outer helices, and an extracellular cap domain which is sequentially between the two pore lining helices of each subunit (Guo and MacKinnon, 2017). Structural studies of the mPIEZO1 channel have been performed using cryo-EM (Ge *et al*., 2015; Guo and MacKinnon, 2017; Saotome *et al*., 2018; Zhao *et al*., 2018; Geng *et al*., 2020; Yang *et al*., 2022) and have resulted in many insights into the arrangement and function of different structural domains of the protein.

However, many crucial aspects of the protein have not been resolved structurally, and so full understanding of the mPIEZO1 channel is lacking. For example, although it is believed that the blade is important for forming the membrane dome around mPIEZO1 and propagating mechanical stimuli to the central pore and thereby triggering opening, the distal end of the blade has not been resolved in any of the mPIEZO1 protein structures (Ge *et al*., 2015; Guo and MacKinnon, 2017; Saotome *et al*., 2018; Zhao *et al*., 2018; Geng *et al*., 2020; Yang *et al*., 2022). Although there are some indications of the structure of the missing three N-terminal HBU, such as the mPIEZO2 structure (Wang *et al*., 2019), which demonstrates the full-length

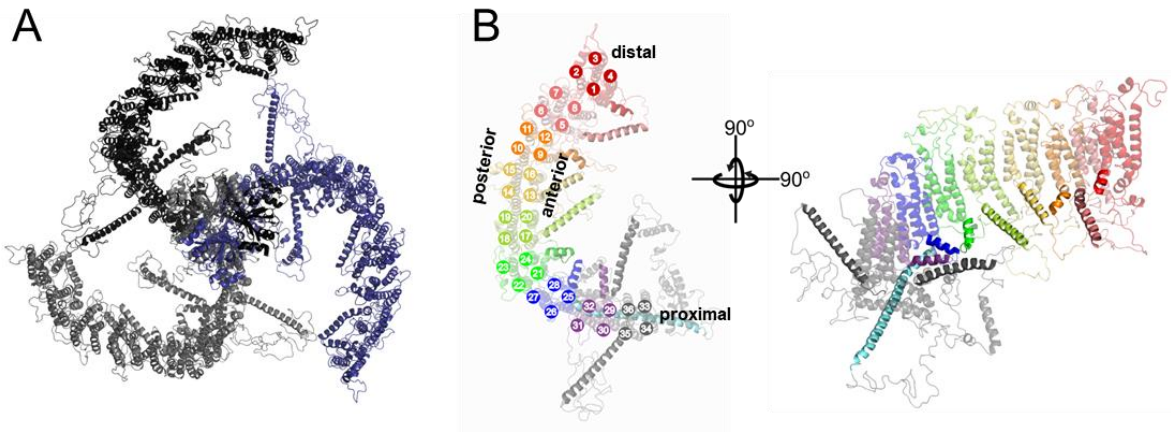


Figure 7.1 The mPIEZO1full-length Blade Structure
The structure of the **A)** trimeric arrangement of the full-length model of the mPIEZO1 protein, with each subunit coloured in black, dark blue and grey, viewed from the extracellular side and **B)** the arrangement of the helical bundles and the helices they contain from the extracellular side and in the membrane plane, where the helices are coloured corresponding to the helical bundle they are part of. To demonstrate the location of the membrane parallel helices, they are shown at full opacity.

blade, the AlphaFold2 model of the mPIEZO1 channel (Jumper *et al*., 2021; Varadi *et al*., 2022), which follows the arrangement of the mPIEZO2 protein blade, and the high sequence homology between the resolved HBUs and those at the distal end of the blade (Chong *et al*., 2021). These have led to the generation of models of the full-length structure, the one used in this work was generated by a former post-doctoral researcher in the Kalli Research Group and has been used for a number of MD simulation studies (Chong *et al*., 2021; De Vecchis *et al*., 2021) (**Fig. 7.1**). The other model that will be discussed in this chapter is the AlphaFold2 model of mPIEZO1 (Jumper *et al*., 2021; Varadi *et al*., 2022), which is highly similar to the full-length model reported in this work. There is only a 9.0 Å RMSD/C_α difference comparing subunit to subunit, which is mostly due to the modelling of MPHs as loops rather than structured helices in the AlphaFold2 model.

Previous MD simulations have identified that the distal HBUs may play a role in the depth and formation of the dome around the mPIEZO1 protein (Chong *et al*., 2021) and the activation of the protein under lateral membrane tension (De Vecchis *et al*., 2021). These studies attempted to understand the differences between the truncated mPIEZO1 protein resolved by cryo-EM (PDB: 6B3R) ($mPIEZO1_{trunc}$) and the full-length mPIEZO1 model ($mPIEZO1_{full\text{-}length}$). However, the number of modelled regions in this model compared to the $mPIEZO1_{trunc}$ structure does not allow clear delineation of the roles of each HBU, the MPHs or the unresolved loop regions. Therefore, I decided to apply a rational truncation approach, in which each HBU of the $mPIEZO1_{full\text{-}length}$ model was removed sequentially and simulated at CG resolution to understand the role of each HBU. This was carried out with the aim to gain structural and functional understanding of dome formation and modulation by the blade, blade structure and dynamics in the Z and XY planes and generation of hypothetical mechanisms for the function of the entire blade in mPIEZO1 activation.

7.2 Results

7.2.1 The Structure of $mPIEZO1_{full\text{-}length}$ Blade Following Simulation

Although this full-length mPIEZO1 structural model has been simulated previously (Chong *et al*., 2021; De Vecchis *et al*., 2021), the effects of simulation on the modelled structure have not been fully investigated and reported. Taking the final snapshot of a 3 μs CG simulation following conversion to atomistic representation and comparing each subunit individually demonstrated that the simulated structure differed in a number of places to the original model made by Chong and colleagues (2021), which contributed to a 9.21 Å RMSD (**Fig. 7.2**).

The arrangement of the HBUs and their TMHs is altered during simulation. The most dramatic change was the angle of the distal three HBUs, which changed from “pointing upwards” in the

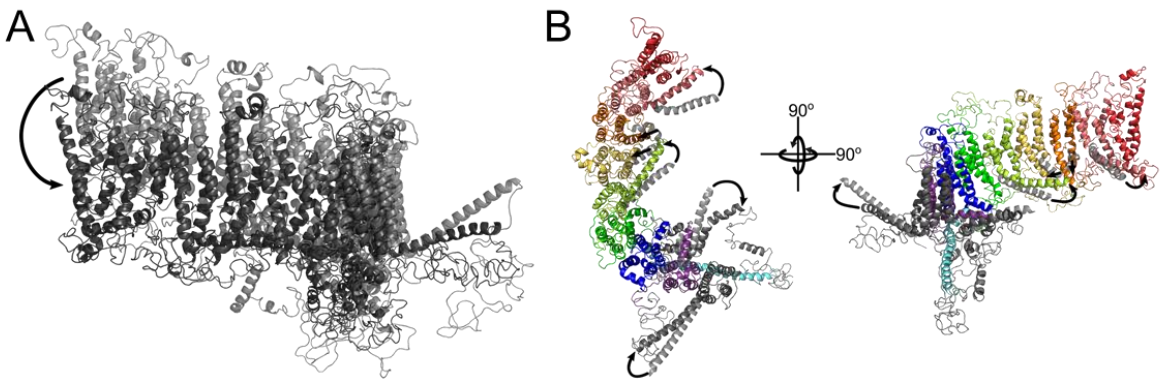


Figure 7.2 The mPIEZO1full-length Protein Following Simulation
The structure of **A)** a single subunit blade viewed from the posterior side in the membrane plane, where the structure prior to simulation is coloured in light grey and the structure following 3 μs of CG simulation and conversion to atomistic representation is in dark grey. The changes to **B)** a single subunit viewed from the extracellular side and from the membrane plane. The membrane parallel helices from pre simulation are shown as transparent light grey helices, and arrows demonstrate their motion during simulation.

membrane (i.e., having an acute angle against the Z axis) to "pointing downwards", the implications of this are discussed later in this chapter, but was suggested by Guo and MacKinnon (2017). A number of changes occurred to the MPHs, where prior to the simulation the model used in this work there appeared to be one MPH corresponding to each HBU, all of which project away from the HBU and towards the posterior of the next blade (Chong *et al*., 2021). Following simulation, however, there were changes to the MPH from HBU1, 3 and 4, which straighten and move to sit beneath the HBU each corresponds to, similar to the arrangement of the beam in the pore-proximal region of the protein. Additionally, the MPH from HBU5 changes angle to sit almost parallel to HBU 3, 4 and 5. This may explain the shift of the MPHs from HBU 3 and 4, as they have to accommodate this positioning of the MPH from HBU5. The other MPH that changes substantially is the posterior MPH from HBU8. This helix is not fully present in the cryo-EM structures of mPIEZO1 other than a short 4 helix turn projection (Guo and MacKinnon, 2017). In the model before simulation, the predicted portion follows the initial trajectory of this short structurally resolved helix, and the following 6 helix turns project upwards towards the extracellular surface. However, upon simulation this MPH adopted a less steep angle against the Z axis, from 47° prior to simulation to primarily 75°, but ranging between 60° and 89° during simulation.

7.2.2 mPIEZO1 Blade Length Alters Membrane Dome Depth

The $mPIEZO1_{full\text{-}length}$ model as described by Chong and colleagues (2021) was truncated by removal of the residues of each helical bundle in turn from the N-terminus to the central pore, to create the protein models $mPIEZO1_{\Delta HBU1}$ to $mPIEZO1_{\Delta HBU1\text{-}8}$, which were individually inserted into an asymmetric model endothelial bilayer (upper leaflet: POPC (55%), DPSM (5%), POPE (20%), cholesterol (20%), lower leaflet: POPC (50%), POPE (20%), POPS (5%), cholesterol (20%), PIP_2 (5%)) (Murphy *et al*., 1992). These systems were then equilibrated for

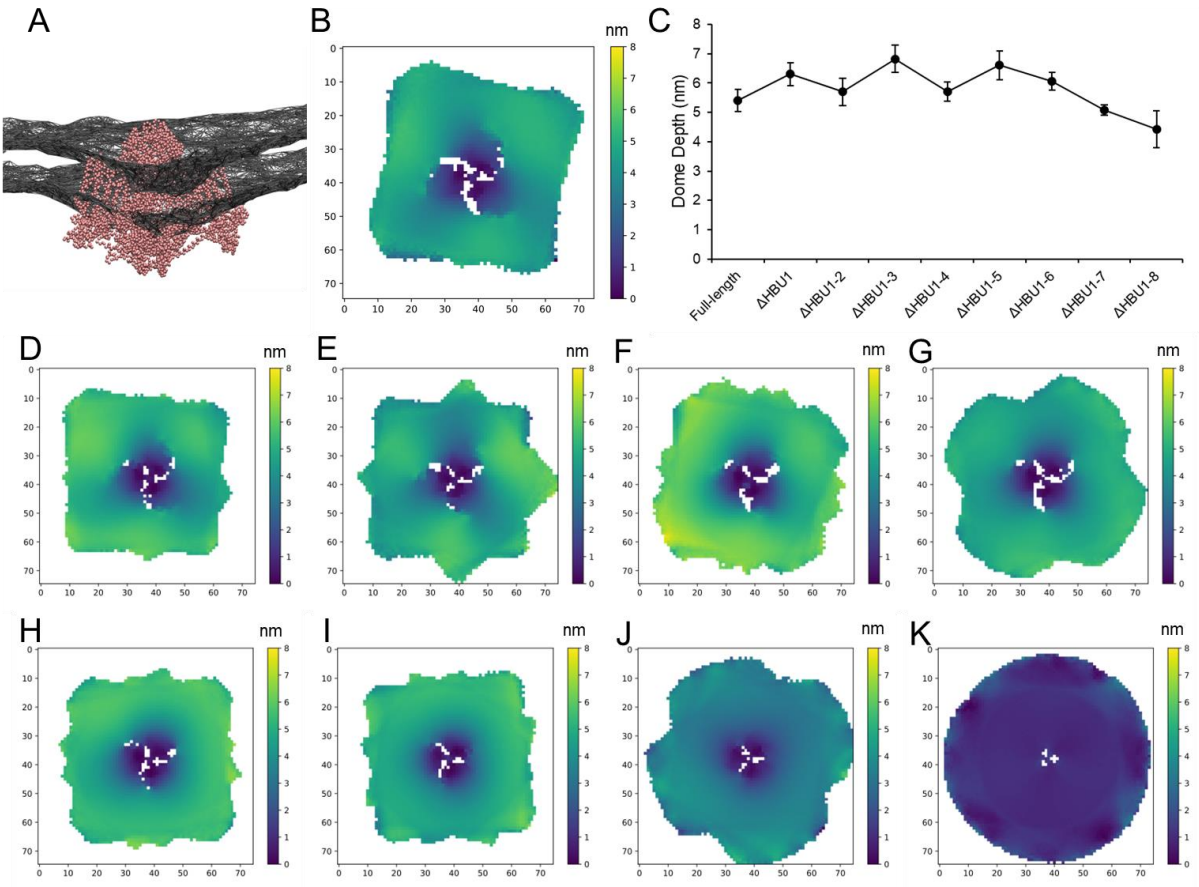


Figure 7.34Truncation of the mPIEZO1 Blade Alters the Depth of the Membrane Dome
A) The membrane dome surrounding the mPIEZO1$_{\Delta HBU1\text{-}6}$ shown in light pink, viewed from the membrane plane, where the PO4 particles of the lipids are shown as grey dynamic bonds. **B)** The height map of the PO4 particles of the membrane around the mPIEZO1$_{full\text{-}length}$ system following averaging of replicates. **C)** The depth of the membrane domes formed around the full-length and truncated mPIEZO1 proteins calculated from the difference in height between the top 20% of heights and bottom 1% and averaged between the lower and upper heightmaps, where the error bars represent standard deviation (n=3). **D-K)** The height maps of each of the truncated mPIEZO1 systems from mPIEZO1$_{\Delta HBU1}$ to mPIEZO1$_{\Delta HBU1\text{-}8}$.

500 ns to allow membrane dome formation. One of the major differences in these systems became apparent at this point: the length of the blade altered the depth of the dome that was formed. These systems were simulated for 3 μs and the height map of the membranes calculated following concatenation and fitting to the protein of the trajectories and averaging across repeats (**Fig. 7.3**).

I had hypothesised that decreased blade length would result in the formation of a shallower dome, as in the previous work by Chong and colleagues (2021). However, the truncated mPIEZO1 models in this work demonstrated a nonlinear depth trend, much like that seen with varying cholesterol concentration (Chong *et al*., 2021). The mPIEZO1$_{full\text{-}length}$ model produces a membrane dome with depth 5.4 nm (note that where one number is reported it is an average of the upper and lower leaflet depth). These numerical values are not directly comparable with the results of Chong and colleagues (2021), as when the heightmap analysis was performed using the top 10% of membrane heights compared to the single lowest most depth as in the

previous study, it resulted in inaccurate results at the lower depths. For example, calculating a depth of -6.24 nm for the upper leaflet dome in the mPIEZO1$_{\Delta HBU1-8}$ system, perhaps due to outliers skewing the data. Therefore, in this work the depth calculation was altered to use the top 20% of Z coordinates compared to the bottom 1%.

The deepest dome depth was reached upon removal of the first three HBUs, where it reaches a maximum of 6.8 nm. Following the loss of HBU5 the change in depth followed a linear trend from 6.61 nm to 4.43 nm in the "flattened" mPIEZO1$_{\Delta HBU1-8}$ system (R^2 = 0.988). The mPIEZO1$_{\Delta HBU1-2}$ and mPIEZO1$_{\Delta HBU1-4}$ systems appear to deviate from the trend, as both have shallower domes than might be predicted from the results of the other systems. This convoluted trend suggested specific roles in membrane dome formation and modulation by 3 different sections of blade: HBUs 1 to 3, HBUs 4 to 6 and HBUs 7 to 9.

7.2.2.1 Each Helical Bundle Plays a Role in Modulating Dome Depth

To understand how these three sections of the blade and each HBU within them modulated the membrane dome depth, the angle against the Z axis over time was measured for each collection of HBUs (HBU1-3, HBU3-6 and HBU6-9 (note that in truncated structures these selections were altered to accommodate the loss of HBUs, for example in the mPIEZO1$_{\Delta HBU1-4}$ system, the selections HBU5-6 and HBU6-9 were used), and histograms generated to demonstrate the spread of angles for each of these selections (**Fig. 7.4**). This indicated that the main component of the blade controlling the depth are the three helical bundles most proximal to the pore (HBU6-9). Their angle against the Z axis ranged from 62° to 77°, but primarily occupied a 70° angle in the WT system, but upon increased dome depth, as in the mPIEZO1$_{\Delta HBU1-3}$ system, the distribution of the blade angle shifted towards a peak at 66° (with a range between 54° and 83°), thereby making the angle more acute and causing the membrane to compensate by forming a steeper and therefore deeper dome. In contrast, in the systems where the depth decreased, such as mPIEZO1$_{\Delta HBU1-7}$ and mPIEZO1$_{\Delta HBU1-8}$, the angle against the Z axis of these HBUs shifted in the opposite direction to a peak of 87° and ranging from 67° to 101° in the case of the mPIEZO1$_{\Delta HBU1-7}$ system. This is demonstrated that the angle against Z of the proximal HBUs is closely associated with the depth of the membrane dome.

Although, the proximal HBUs can be more easily linked to the depth of the dome, it was changes to the distal HBUs that result in the deepest membrane dome. However, the angle against Z of the distal three HBUs, were somewhat harder to link to the membrane depth, as this could only be measured for 2 of the systems in this investigation. Despite this, the angle of HBU1-3 in the mPIEZO1$_{full-length}$ and HBU2-3 in the mPIEZO1$_{\Delta HBU1}$ provided the basis for understanding increased depth upon loss of the distal HBUs. In these systems, these HBUs

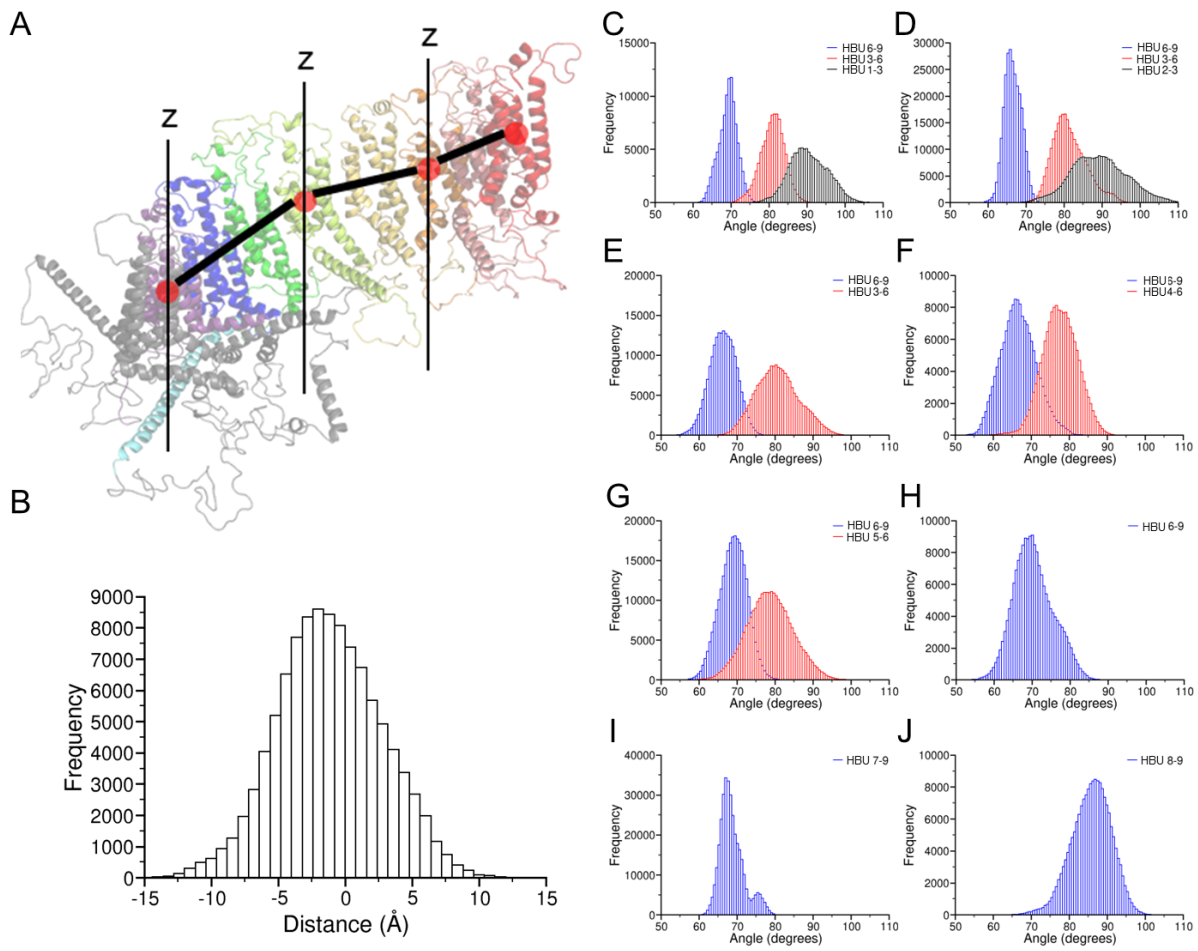


Figure 7.4 The Contribution of Different Regions of the Blade to the Membrane Dome Depth
A) The selections of the blade used to measure the angle against the Z axis of different regions of the blade. **B)** The frequency histogram demonstrating the Z distance between the centre of geometry of HBU4 and HBU1. The frequency histograms of the angles defined in panel A from the **C)** mPIEZO1$_{\text{full-length}}$, **D)** mPIEZO1$_{\Delta\text{HBU1}}$, **E)** mPIEZO1$_{\Delta\text{HBU1-2}}$, **F)** mPIEZO1$_{\Delta\text{HBU1-3}}$, **G)** mPIEZO1$_{\Delta\text{HBU1-4}}$, **H)** mPIEZO1$_{\Delta\text{HBU1-5}}$, **I)** mPIEZO1$_{\Delta\text{HBU1-7}}$, **J)** mPIEZO1$_{\Delta\text{HBU1-8}}$, following concatenation of the replicates and each of the chains.

adopt angles between 77° and 105°, or 69° to 110°, respectively. This indicates their high level of flexibility in the Z plane, but also angles larger than 90° indicate that the vector is pointing towards the intracellular side of the membrane, rather than the extracellular side as seen in the starting structure. To assess how significant this downward direction is in terms of the mPIEZO1$_{\text{full-length}}$ protein blade structure, the distance between the centre of geometry of HBU 1 and HBU 4 in the Z direction over time was measured. The resulting distance was treated as the angles were and converted to a histogram to best visualise and understand the range of distances adopted by this region of the blade. This demonstrated that the first three HBUs point downward in the membrane the majority of the time, resulting in HBU1 being up to 14 Å lower than HBU4, but most frequently 2 Å lower in the Z direction. This is particularly striking as it explains how loss of these HBUs results in a deeper dome, as they are acting to reduce the depth of the dome by pointing downward rather than upward. Loss of these appears to allow the protein to “sink” into the membrane, thereby making the dome deeper.

The downward angle of the distal three HBUs also explains the complex shape of the dome formed by mPIEZO1. The height map of the mPIEZO1$_{\text{full-length}}$ system demonstrated the trilobed topology of the dome that radiates outward from the protein. The direction of the three distal HBUs in the Z direction can explain how the depth is maintained between the blades of the protein and beyond the channel itself, as the membrane beyond the distal HBUs has to bend to accommodate the curvature of the protein, not only around the pore but along the entire blade. As the blade was truncated and the downward bend of the distal HBUs was reduced, the extension of the dome around the channel also reduced. This is seen upon the loss of the three distal HBUs in the mPIEZO1$_{\Delta\text{HBU1-3}}$ system, although the extended footprint was never fully abolished until loss of the dome entirely in the mPIEZO1$_{\Delta\text{HBU1-8}}$ system.

The main change to the angle of HBUs 3 to 6 across the different systems was not a shift to more acute or obtuse angles, but an increase in flexibility of this region upon truncation of the distal HBUs. In the mPIEZO1$_{\text{full-length}}$ system, the angle of this region was between 71° and 90°, only slightly wider than that of the angle HBU6 to 9, but upon loss of N-terminal HBUs this became more apparent, as it increased to 71° to 97° upon loss of HBU1 and to 66° to 99° following loss of HBU2. However, in the mPIEZO1$_{\Delta\text{HBU1-3}}$ system, the angle range reduced to resemble the mPIEZO1$_{\text{full-length}}$ system, albeit with a slight shift to acuter angles, peaking at 76° and ranging from 60° to 91°. The loss of HBU4 again demonstrated an increase in range of the angle as it can adopt angles between 60° and 98°. This follows the trend of the depth of the membrane dome, and it appears that when this region of the blade is able to adopt larger Z angles due to its flexibility, the dome depth decreases. This appears to act as a

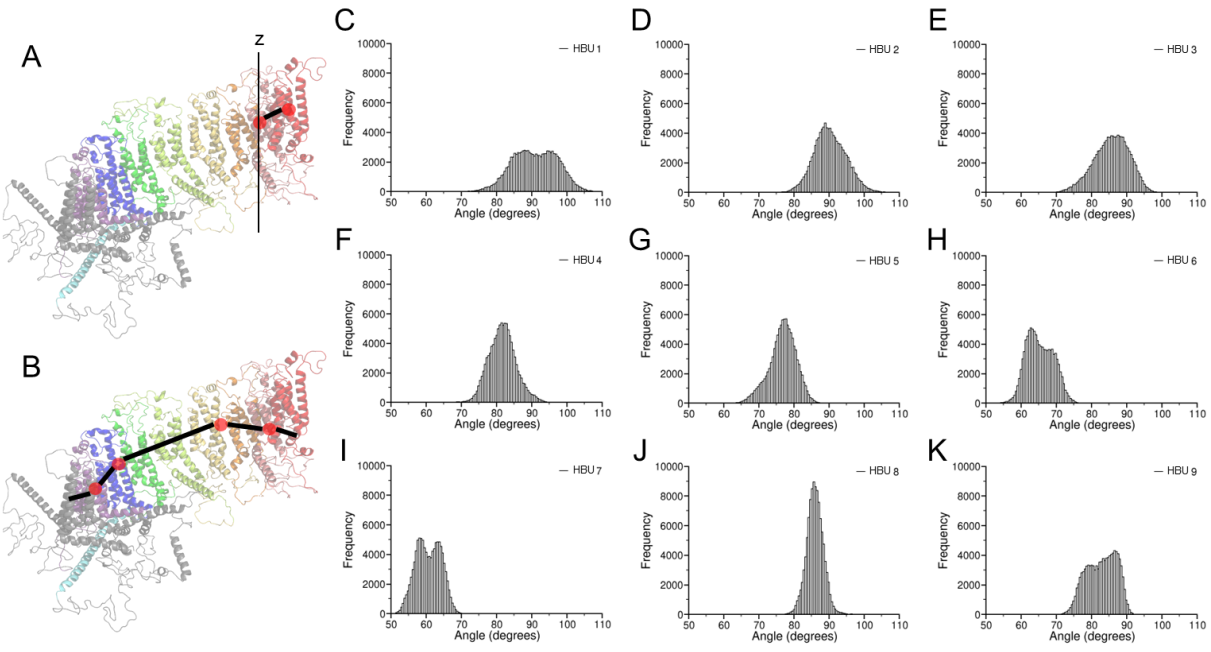


Figure 7.5 The Flexibility of each of the Helical Bundle Units
A) An illustration of the selections used to measure each of the Z angles of the HBUs and **B)** the points of flexibility along the blade represented as red points. **C-K)** The frequency histogram generated following concatenation of the Z angle of the HBU between replicates and chains for each of the HBUs.

compensatory mechanism for the loss of the distal HBUs and their downward slant, thereby maintain the dome depth in the mPIEZO1$_{\Delta HBU1-2}$ and mPIEZO1$_{\Delta HBU1-4}$ systems, although this did not occur in the mPIEZO1$_{\Delta HBU1-3}$ simulations.

The variability of HBU 3 to 6 angle and its reduced flexibility in the mPIEZO1$_{\Delta HBU1-3}$ system indicated that a more granular approach was needed to understand the role of each individual HBU in controlling the Z angle of the blade and the resulting changes to the membrane dome. Therefore, the angle generated from the vector between the centre of geometry of one HBU and the next was measured for the mPIEZO1$_{full-length}$ simulations (**Fig. 7.5**). This revealed that the blade is composed of several regions of differing flexibility that contribute to its ability to adopt different angles against the Z axis. The main flexible regions were HBU 1, which could adopt angles from 73° to 107°, and HBU3, which ranged from 69° to 98°. At the proximal end of the blade, HBU 6 and 7 demonstrated that they each seemed to be able to adopt two independent angles in the Z axis, of 63° and 69° for HBU6 and 58° and 64° for HBU7.

The flexibility and angle of each of the HBUs in the Z plane explains the increase and subsequent decrease in depth upon truncation. Following the loss of the first two HBUs, the 3 to 5 region of the blade is able to compensate and increase its angle against the Z axis, thereby flattening and maintaining the depth below 6 nm. However, loss of HBU3 prevents this region of the blade from compensating, as the next flexible HBU is HBU6, thereby causing the dome to become deeper as a result. Loss of HBU3 allows some further flexibility of the blade before HBU6, but loss of these HBUs results in the loss of membrane dome depth. This can be explained by the angle of each of the HBUs, as HBU8 and 9 only point “upwards” by 12° to 18° and so do not substantially contribute to the angle of the blade and membrane dome formation. Therefore, in the full-length protein the first three HBUs of the blade act to moderate the depth and prevent the protein from “sinking” into the bilayer, while HBU6 and 7 contribute the most to the upward angle of the protein blade and the subsequent formation of the membrane dome.

7.2.3 mPIEZO1 Blade Length Alters Protein Dynamics

Following the identification of dynamic changes in the blade in the Z axis, the overall dynamics of the protein structure was investigated to see how this might be altered by truncation of the blade (**Fig. 7.6**). Initially, the RMSD/C_α and RMSF/C_α of the protein was measured, which demonstrated that the distal regions of the blades and the intracellular loops are more flexible than the rest of the structure, and the overall RMSD/C_α of the proteins demonstrated that flexibility of the structure was maintained until loss of all 8 HBUs, where the protein became very stable**.** However, the RMSD/C_α and RMSF/C_α values included movement in the Z plane, which may have been disguising other changes in dynamics.

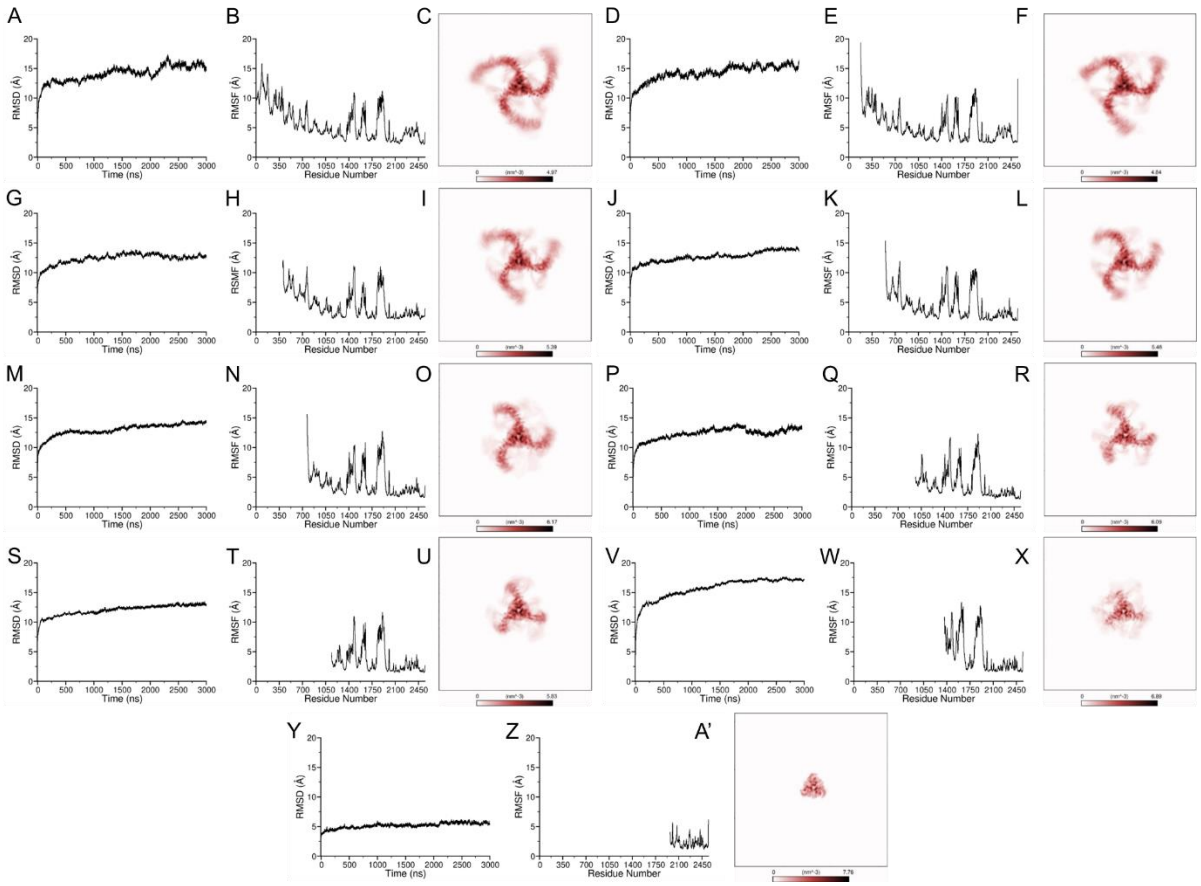


Figure 7.6 The Dynamics of the Full-length and Truncated mPIEZO1 Proteins
The **A)** RMSD/C_α of the backbone atoms of the mPIEZO1$_{full\text{-}length}$ protein and **B)** RMSF/C_α of the mPIEZO1$_{full\text{-}length}$ protein averaged between repeats and subunits, and the **C)** 2D protein density of the backbone atoms of the mPIEZO1$_{full\text{-}length}$ protein following fitting of the simulation to the structure and averaging between repeats. The dynamics of the **D-F)** mPIEZO1$_{\Delta HBU1}$, **G-I)** mPIEZO1$_{\Delta HBU1\text{-}2}$, **J-L)** mPIEZO1$_{\Delta HBU1\text{-}3}$, **M-O)** mPIEZO1$_{\Delta HBU1\text{-}4}$, **P-R)** mPIEZO1$_{\Delta HBU1\text{-}5}$, **S-U)** mPIEZO1$_{\Delta HBU1\text{-}6}$, **V-X)** mPIEZO1$_{\Delta HBU1\text{-}7}$, **Y-A')** mPIEZO1$_{\Delta HBU1\text{-}8}$.

To investigate the dynamics in the XY plane, the density of the backbone particles of the proteins were plotted. The mPIEZO1$_{full\text{-}length}$ blade dynamics mirrored previous CG simulations with this protein, as it demonstrated the reduced density of the N-terminal of the blades and their heightened flexibility compared to the central regions of the structure (Chong *et al*., 2021). As the blade was truncated, the tip of the blade remained flexible in the XY plane, but the pore and proximal MPHs became more defined in the density, indicating they became more rigid. Upon comparison of the plots between truncations, it was noted that the curvature of the blades appeared to change upon truncation of the distal helical bundles. Therefore, three-point angle analysis was used to measure the angle between the centre of geometry of different HBU selections (**Fig. 7.7**). Initially, the angle between HBU9-6-1 of the mPIEZO1$_{full\text{-}length}$ system was measured and each chain and replica combined to produce a frequency histogram. This demonstrated that the blade is able to adopt angles between 82° and 124°, with two main populated conformations at 103° and 112°. However, summarising 5 HBUs as a single vector may have resulted in the loss of information and states adopted by the mPIEZO1$_{full\text{-}length}$ blade.

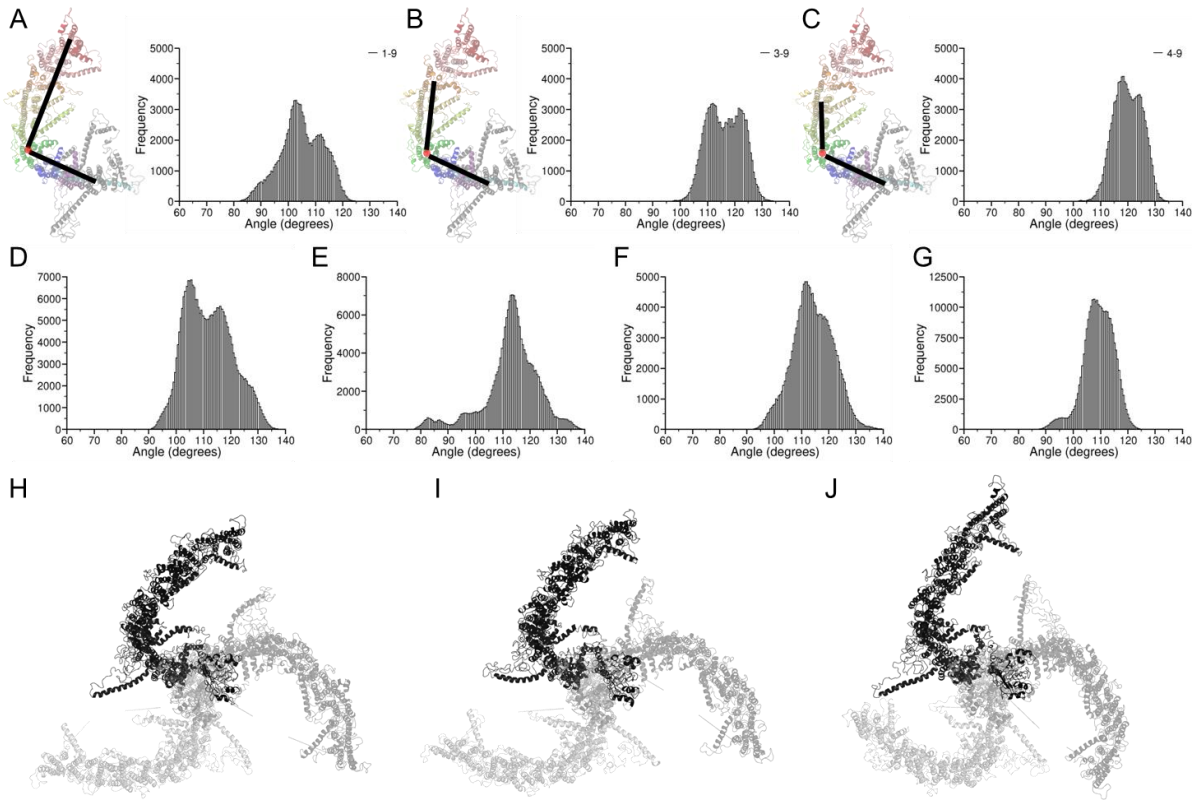


Figure 7.7 The XY Dynamics of the mPIEZO1 Blade
The selections of HBUs for measuring the XY angle of the blade using **A)** the HBU1-6-9, **B)** HBU3-6-9 and **C)** HBU4-6-9 angles alongside the associated frequency histogram. The angles generated by measuring the three-point angle of HBU3-6-9 in the **D)** mPIEZO1$_{\Delta HBU1}$, **E)** mPIEZO1$_{\Delta HBU1-2}$, **F)** mPIEZO1$_{\Delta HBU1-3}$ and **G)** mPIEZO1$_{\Delta HBU1-4}$. The blade structures of mPIEZO1$_{full-length}$ where chain A (black cartoon) is at **H)** 112°, **I)** 118° and **J)** 122° during CG simulation following conversion to an atomistic representation for clarity the other two chains are represented as transparent cartoon.

To assess whether a better selection of HBU was available, the angle of each of the helical bundles in the XY plane was assessed to identify potential points of flexibility (**Fig. 7.8**). These were measured from the centre of geometry of the next helical bundle to the centre of geometry of the posterior proximal helix to the centre of geometry to the anterior distal helix in the HBU of interest. This resulted in the identification of three points of flexibility in the XY plane of the blade: At HBU6, HBU4, and HBU3. Therefore, the angles between HBU9-6-4 and HBU9-6-3 were measured to see if this provided more information about the dynamics of the mPIEZO1$_{full-length}$ blade during CG simulation. This resulted in the shift of the identified conformational angles, but importantly this revealed that there are three main blade conformations in the XY plane.

The angle between HBU9, HBU6 and HBU3 (or the distal most HBU in systems where the third HBU had been removed) was selected for measuring the blade angle in the XY plane, as it provided the most distinction between the three states and so allowed better comparison between systems. Measuring this angle demonstrated that the WT protein adopts three main blade angle conformations, at 112°, 118° and 122°, but these are very flexible with adoption

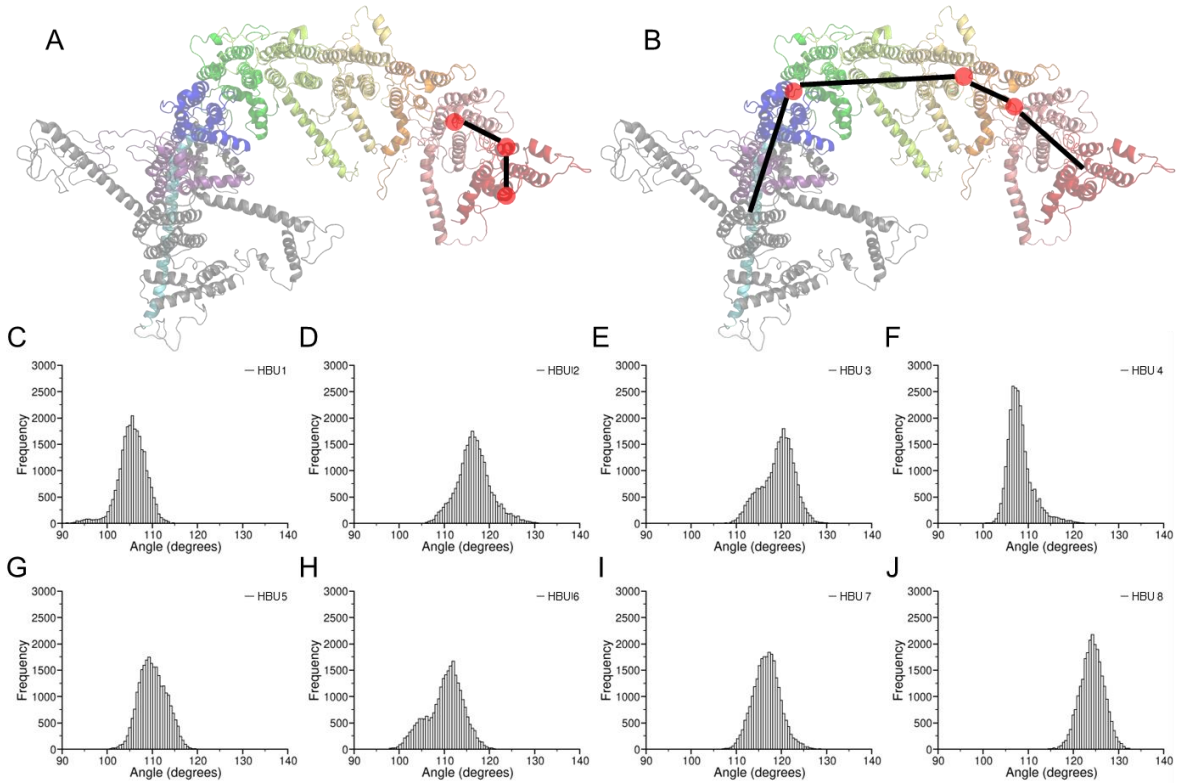


Figure 7.8 The Flexibility of the HBUs of the Blade in the XY Plane
A) The selections of HBUs for measuring the XY angle of the HBUs and **B)** the points of flexibility along the blade represented as red points. **C-J)** The frequency histograms of the angles adopted by each of the HBUs in the mPIEZO1$_{\text{full-length}}$ system.

of angles between these main peaks. The frames corresponding to the 112°, 118° and 122° angles of chain A were separately extracted from the trajectory of the 3 replicas and used to identify the blade structure at these angles. The 112° angle is a compact structure, comparable to the initial starting position of the protein, where each of the flexible regions of the blade (HBU3 and HBU6) are at acute angles. Whereas the 118° angle is a slightly more relaxed conformation, where the angle of HBU6 has increased. The final 122° angle is a very open blade conformation which is exemplified by the straightening of the first three HBUs to follow the trajectory of HBU4 and HBU5.

Truncation of the distal HBUs had a dramatic reductive effect on the flexibility of the blade. Initially, loss of the first HBU resulted in loss of distinction of the three states seen through measurement of the HBU9-6-3 angle, but truncation of HBU2 caused there to be only one main populated state at 113°, which was maintained upon further truncation. The blade was able to still adopt more acute and obtuse angles, but this was increasingly eliminated as the blade was truncated. Comparison of the 113° blade structure of mPIEZO1$_{\Delta\text{HBU1-2}}$, where this is the predominant state, and resembles the 112° blade structure of the mPIEZO1$_{\text{full-length}}$ blade system. Thereby implying a role for the distal HBUs in allowing mPIEZO1 to occupy three distinct blade conformations.

7.2.4 Structural Understanding of the Flexible Regions of the mPIEZO1 Blade

It was initially unclear what was responsible for allowing some regions of the blade to be flexible in the Z or XY planes compared to others, so I performed protein contact analysis between each HBU and its neighbour to identify the interactions between them (**Fig. 7.9**). By defining the interface between the HBUs and how this might change upon change of the Z angle, I hoped to identify a mode of motion between the HBUs, such as the motions between domains in the tilting as in the major facilitator super family (Quistgaard *et al*., 2016) or elevator-like motions, as in the glutamate transporter family (Pavić *et al*., 2019). This analysis revealed that the interface between HBUs during CG simulation was not very extensive, ranging between 5 to 20 interacting residues for each HBU. The most extensive interface was between HBU 1 and HBU 2, whereas HBU 5 and 6 were only interacting *via* 5 residues.

Despite the low number of contacts between the HBUs, I attempted to identify any differences in the contacts between HBUs at different angles by extracting the simulation frames with either 112°, 118° or 122° angles of the chain A blade, which were used to identify the contacts in each state. This did not identify any significant changes in the interactions when adopting different XY angles. However, the minimal interface between HBUs does explain the high level of flexibility of the blade, as it is able to move without disrupting very strong or extensive inter-HBU interactions.

Due to there not being clear changes between the HBUs, I investigated whether a change in angle was occurring within the HBUs themselves. Therefore, the Z tilt angle of each helix in

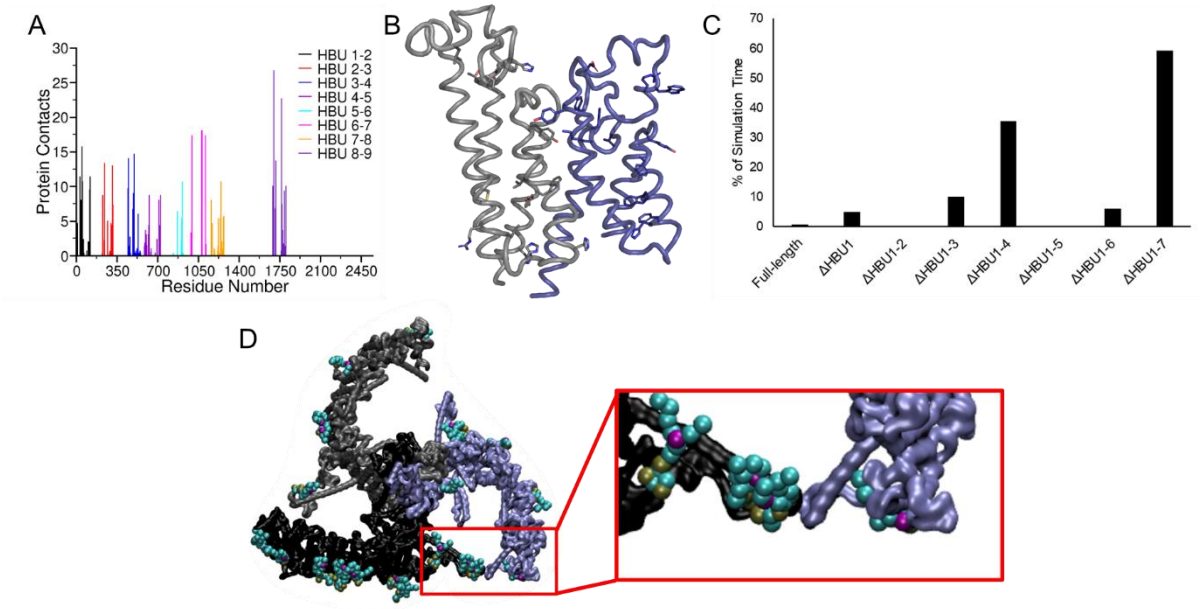


Figure 7.9 The Interfaces Between the Helical Bundles and the Membrane Parallel Helices
A) The contacts between each HBU and the adjoining HBU and **B)** a representation of the residues shown as sidechain sticks forming contacts between HBU1 in grey and HBU2 in dark blue. **C)** The interactions between the posterior MPH from Chain A and the anterior MPH from Chain C represented as the % of simulation time the distance between the MPHs is less than 1.5 nm. **D)** The bridging of the MPHs from two subunits (QuikSurf representation in VMD in black and ice blue) by PIP_2 lipids shown as VDW representation.

chain A was measured over the 3 mPIEZO1$_{full\text{-}length}$ repeats. For the most part, there was no considerable difference in the angles, other than those that would be anticipated by the tilting of the entire HBU. The helices in adjoining HBUs also did not demonstrate any clear differences when the flexible HBUs moved. Having mostly eliminated the possibility of interactions in the helical bundles themselves, I examined the structural features around the points of flexibility or rigidity.

A major structural feature that differed for each of these regions was the presence of MPHs. Each of the stable regions were supported by a MPH, either below or lateral to the HBUs. During simulation, the mPIEZO1$_{full\text{-}length}$ blade structure underwent some structural changes of the modelled regions, during which the MPH from HBU1, 3 and 4 moved from pointing laterally away from the HBU to straightening and either sitting beneath the HBU it corresponds to as for HBU1 and HBU4, or extending the final helix of the HBU, as for HBU3. The MPH from HBU5 also changed angle to be alongside HBUs 3, 4 and 5. These rearrangements meant that the stable regions of the protein were in direct contact with a MPH or the beam, whereas the flexible HBUs, in both the Z and XY directions, were out of range of an MPH, as in the case of HBU3 and HBU6, each of which spatially precede and follow a MPH.

7.2.4.1 The Role of the Membrane Parallel Helices in mPIEZO1 Dynamics

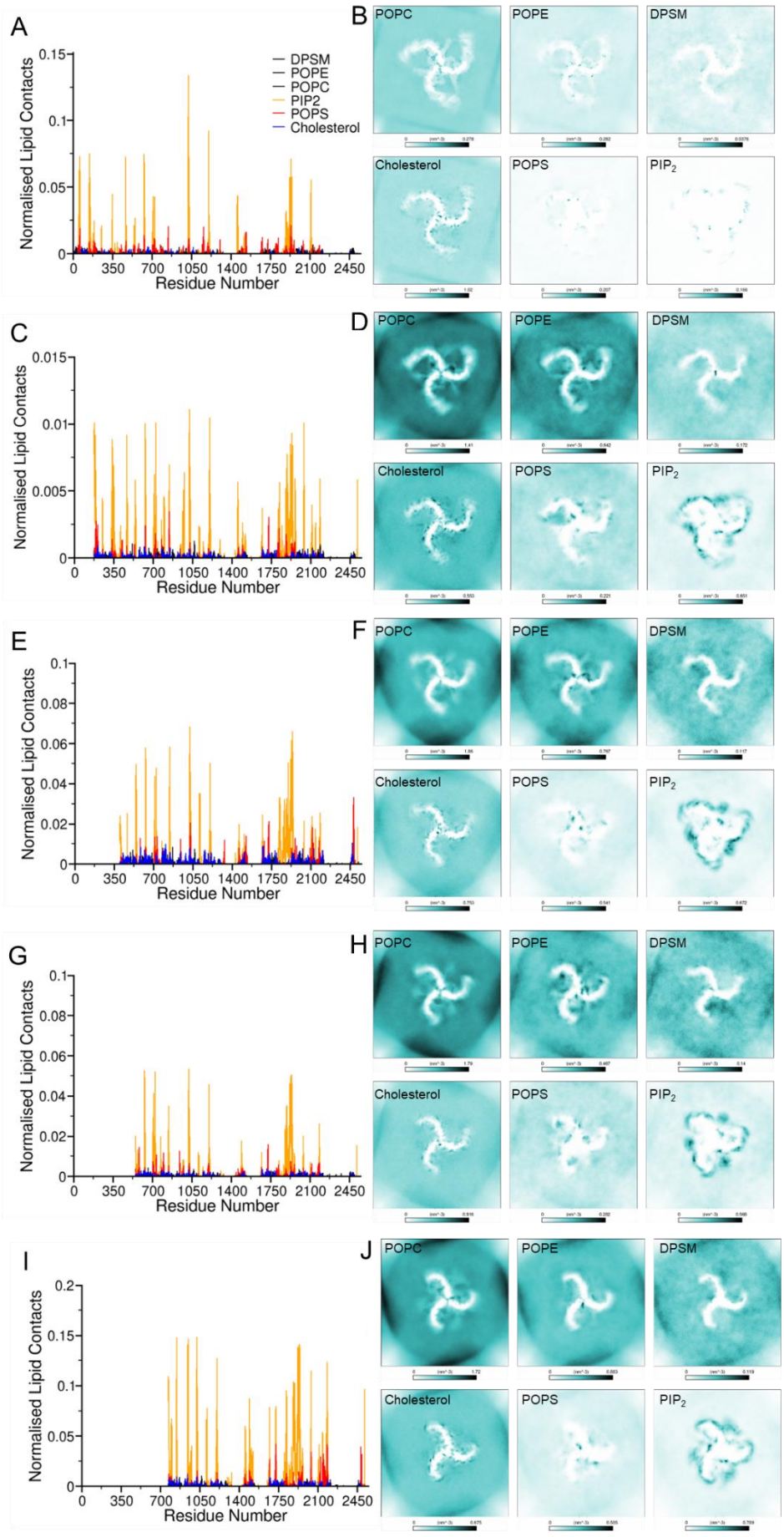
The identification of MPHs as structural features differentiating the rigid and flexible regions of the blade caused attention to be paid to the angles they adopt during the course of the simulations. The Z angle was measured by selecting the centre of geometry of the first and last helical turns of each of the MPHs from HBU2, 4, 5, 8 and 9, where possible due to truncations, for each of the mPIEZO1 systems and following concatenation of the replicas and chains, frequency histograms were produced. This demonstrated the high stability of each of the MPHs, as they rarely altered their angle during simulation or following truncation. With the exception of the MPH from THU5, which shifts from most frequently occupying an angle of 102° in the Z plane in the mPIEZO1$_{full\text{-}length}$ system, to 113° in the mPIEZO1$_{\Delta HBU1\text{-}3}$ system. This increase in angle corresponds to the MPH having a steeper upward angle in the bilayer, as the furthest point of the helix is “above” the helix turn closest to HBU5 and so as it becomes steeper upwards the angle against the Z plane increases. This corresponded to the increased depth, as it had to accommodate the curvature of a deeper dome and increase its angle. Additionally, in the shallowest dome, created by mPIEZO1$_{\Delta HBU1\text{-}7}$, the anterior MPH from HBU8 adopts a wider angle in the Z plane, also corresponding to having to accommodate a reduced curvature in the bilayer.

The angle of the MPH was measured in the XY plane using a three-point angle, from the centre of geometry of the outer pore helix to the centre of geometry of the first helical turn to

the centre of geometry of the last helical turn of the MPH. These data showed that the MPH are slightly more flexible in the XY plane, but only the MPH from HBU5 and both from HBU8 demonstrate more than one main conformation in this plane in the mPIEZO1$_{full\text{-}length}$ system. However, upon loss of the distal HBUs, these states become less distinct. In the mPIEZO1$_{\Delta HBU1\text{-}2}$ system, these MPHs adopt one main angle, which is at a comparable angle to one of the mPIEZO1$_{full\text{-}length}$ protein states. When the first three HBUs were lost, there were a number of changes to the MPH angles, as both the MPHs from HBU3 and HBU4 adopted more acute angles, suggesting a contraction of the mPIEZO1 blade at these regions. Removal of HBU4 destabilised the MPH from HBU5 considerably, likely because these two regions were alongside each other and so removal of HBU4 increased the range of motion of the MPH from HBU5 in the XY plane. As the blade became more truncated, the angle of the proximal MPHs became more variable, unlike the overall blade angle, which suggested that the distal HBUs or MPHs have a role in stabilising these otherwise flexible proximal regions.

Alongside modulating the flexibility of their own blade, the MPHs also appeared to form interactions between the subunits, as they face towards the posterior of the next blade and are accessible by the posterior MPH from HBU8, and thereby alter the adjoining blade angle. These were not canonical protein-protein interactions, but rather seem to be mediated by interactions with PIP_2 lipids (as discussed later in this chapter). The distance between the MPHs on each chain were measured, and distances below 1.5 nm were considered interactions. This revealed that for mPIEZO1$_{full\text{-}length}$, there appeared to be extensive interactions between the MPHs from HBU2 and the posterior MPH from HBU8 from the next subunit, with this occurring in 60.4% of the simulation time. This arrangement appeared to coincide with the more compact blade angles. Interestingly, the loss of the distal MPHs caused the latter MPHs, such as the anterior MPH from HBU8 to form interactions with the posterior MPH from HBU8 from the neighbouring subunit more regularly, increasing from 0.6% of simulation time in the mPIEZO1$_{full\text{-}length}$ to a maximum of 59.2% of simulation time in the mPIEZO1$_{\Delta HBU1\text{-}7}$ system, thereby maintaining the angle and arrangement of the proximal MPHs, which is lost upon truncation of these stabilising helices.

Throughout the simulations, the MPHs act as stiff beams either alongside or below the HBUs they correspond to or neighbour and act to control the motion of the blade in the XY and Z directions. This prevents large scale movements and bending, as well as maintaining the inward bend of the overall beam through interactions with the other subunits. Maintenance of the arrangement and angle of the nearby HBUs gives rise to points of flexibility along the blade at HBU3 and HBU6, as these are not supported by or in extensive direct contact with a MPH.


A
B
C
D
E
F
G
H
I
J
Normalised Lipid Contacts
Residue Number
DPSM
POPE
POPC
PIP2
POPS
Cholesterol
POPC
POPE
DPSM
Cholesterol
POPS
PIP2

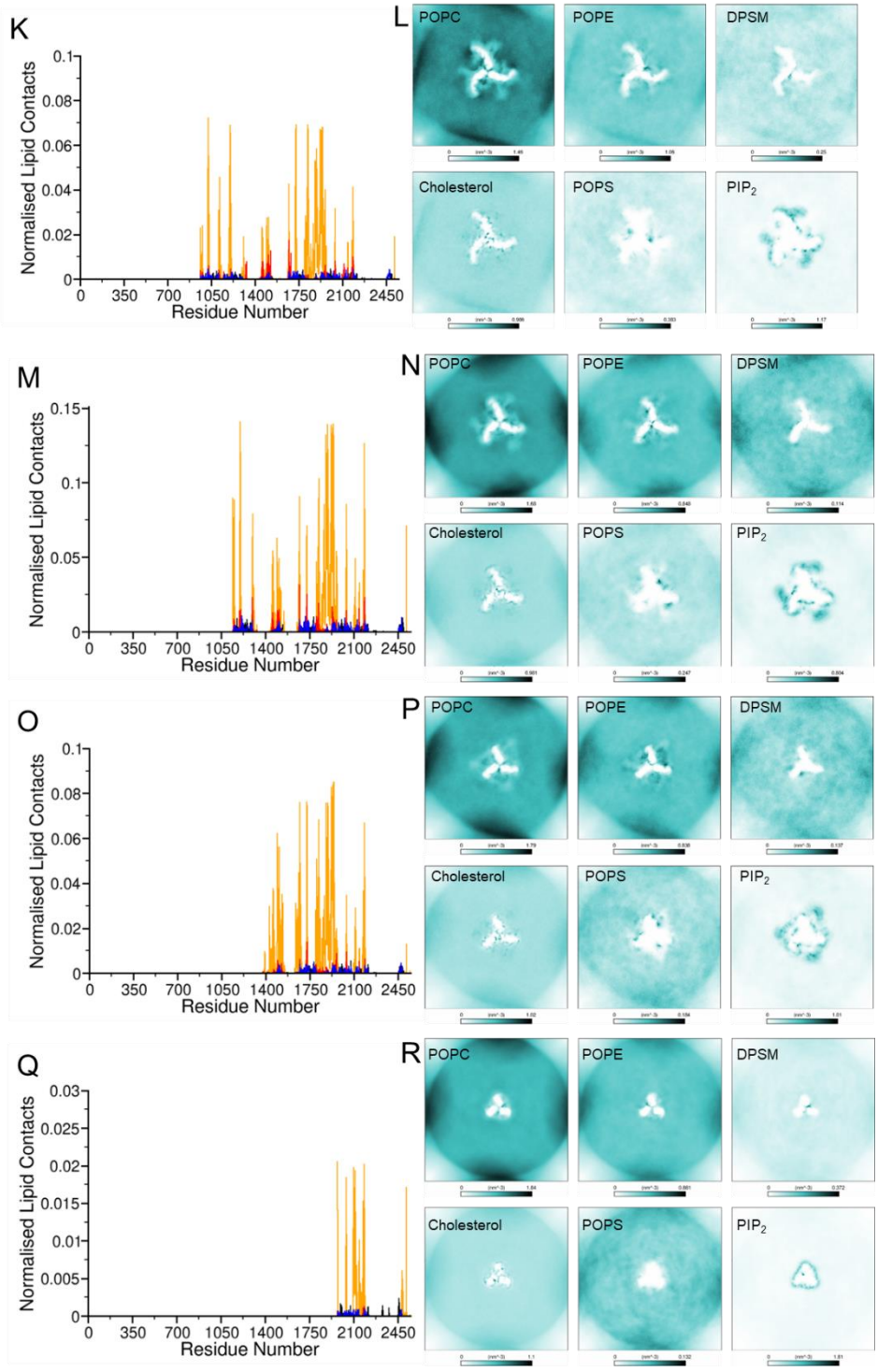


Figure 7.10 Lipid Contacts with the Full-length and Truncated mPIEZO1 Proteins
The contacts between lipids and the protein normalised by the number of frames and lipid in the bilayer and averaged between replicates and chains, and the 2D density of the PO4 atom of the lipids in the bilayer for the **A-B)** mPIEZO1$_{\text{full-length}}$, **C-D)** mPIEZO1$_{\Delta HBU1}$, **E-F)** mPIEZO1$_{\Delta HBU1-2}$, **G-H)** mPIEZO1$_{\Delta HBU1-3}$, **I-J)** mPIEZO1$_{\Delta HBU1-4}$, **K-L)** mPIEZO1$_{\Delta HBU1-5}$, **M-N)** mPIEZO1$_{\Delta HBU1-6}$, **O-P)** mPIEZO1$_{\Delta HBU1-7}$, **Q-R)** mPIEZO1$_{\Delta HBU1-8}$.

7.2.5 Lipid Contacts with Truncated mPIEZO1 Blades

Previous work on mPIEZO1 has primarily focused on the effects of general membrane properties on protein activity (Morachevskaya *et al*., 2007; Chubinskiy-Nadezhdin *et al*., 2011; Conrard *et al*., 2018; Romero *et al*., 2019; Ridone *et al*., 2020), but more recently simulations have allowed the investigation of specific interactions between lipids and the channel (Chong *et al*., 2021)**.** This has led to the identification of a preference for interactions with cholesterol and PIP_2 lipids at specific sites, which may have biological roles. However, previous work has not seen interactions with all of the available CRAC and CARC cholesterol binding sites (Fantini and Barrantes, 2013; Chong *et al*., 2021). Therefore, investigating the lipid distribution and contacts upon loss of the distal HBUs was of particular interest, as less well-defined sites with reduced affinity might appear, or there could be an overall change to the lipid interaction profile (**Fig. 7.10**).

Overall, the profile of cholesterol binding did not change dramatically between the truncated systems, other than slightly increased interaction frequencies in some systems. However, a very striking change was the increased anionic lipid presence at the pore of the mPIEZO1 channels. This was very rare in the $mPIEZO1_{full\text{-}length}$ system, but loss of the first HBU led to interactions with PIP_2, and loss of the second led to increased POPS in the same region. As these lipids were in the cytoplasmic leaflet, this may have been due to the change in MPH arrangement, as the loss of interactions between the MPH of a subunit and the posterior MPH from HBU8 may have allowed more anionic lipids to diffuse into this space.

The lipid contacts and density analysis confirmed that the interactions between the MPHs were anionic lipid mediated, as increased PIP_2 binding at the anterior MPH from HBU8 corresponded to increased contacts with the posterior MPH from HBU8. Interestingly, the number of PIP_2 contacts with the posterior MPH from HBU8 increased upon loss of the distal HBUs and their associated MPHs, not only in terms of number of contacts with residues, but also contacts with other residues on the helix. This indicated that there is competition for PIP_2 lipid binding between the MPHs, and loss of these sites allows increased saturation of the others, as this is also seen for the anterior MPH from HBU8. The presence of PIP_2 binding at the posterior MPH and other proximal MPHs, may also explain how these regions contribute to the depth and formation of the dome, as PIP_2 is an inverse conical lipid (Pomorski *et al*., 2014), and could be encouraging the membrane dome curvature to increase. This may also contribute to the increased curvature on the anterior side of the blade, as it has more MPHs and anionic lipid binding sites, and the expansion of the membrane curvature beyond the protein by HBU1-3, as increased PIP_2 concentration at the MPH from HBU2 could deform the membrane further.

In addition to the lipid contact analysis, 2D lipid density analysis was performed and revealed that there are lipids interacting with the pore between the outer pore TMH. These were typically neutral lipids, either POPC or POPE, but in the mPIEZO1$_{\Delta HBU1}$ system a DPSM lipid interacts with the pore in addition to POPC and POPE. There is some dispute about whether the presence of lipids in the mPIEZO1 pore is real or a simulation artefact (De Vecchis *et al*., 2021; Jiang *et al*., 2021b; Jiang *et al*., 2022), but following a recent structural study demonstrating lipids in the mPIEZO1 pore (Yang *et al*., 2022), these are now considered to be a reliable structural detail. The lipid binding sites identified in this study corresponded to those seen in the structures of mPIEZO1, and so provided some additional reliability that the interactions between mPIEZO1 and the membrane in these simulations are mimicking those *in vitro*.

The lipid density in the pore was maintained upon truncation of mPIEZO1, even in the mPIEZO1$_{\Delta HBU1-8}$ system. This demonstrated that truncation of the blade or change in structure and the resulting alterations to the depth and curvature of the bilayer and interactions of lipids does not significantly affect the pore structure in CG systems. This may be due to the presence of an elastic force network, or the role of the elbow region to ensure that it is mechanical stimuli that propagates changes in the blade structure to the pore, rather than solely a loss or gain of curvature.

7.2.6 The Role of the Beam Region in mPIEZO1 Structure and Function

Much of the data reported here indicate that the MPHs modulate the dynamics of the blade in both the Z and XY planes. Therefore, to investigate this further, the structured element of the beam was removed and replaced with an unstructured loop, to understand what effect this would have on the protein dynamics. The beam was chosen as it is the most well characterised of the MPHs, as it is structurally defined and generally agreed to have a role in activation and propagation of stimuli to the pore (Zhao *et al*., 2018; Wang *et al*., 2018; Zhao *et al*., 2019). The resulting mPIEZO1$_{Beam}$ system, was simulated as before in a model endothelial membrane for 3 μs, to understand how this MPH modulates the protein structure and dynamics.

7.2.6.1 Modulating Dome Depth

Replacement of the structured beam with a loop resulted in reduction in depth of the membrane dome that forms around mPIEZO1, from 5.4 nm to 4.7 nm (**Fig. 7.11**). This was of particular interest, as the angle of the beam against the Z axis, measured from the centre of geometry of the first and last helical bundle, did not change in the truncated systems in relation to their depth. However, the reduced depth was comparable to the mPIEZO1$_{\Delta HBU1-7}$ system,

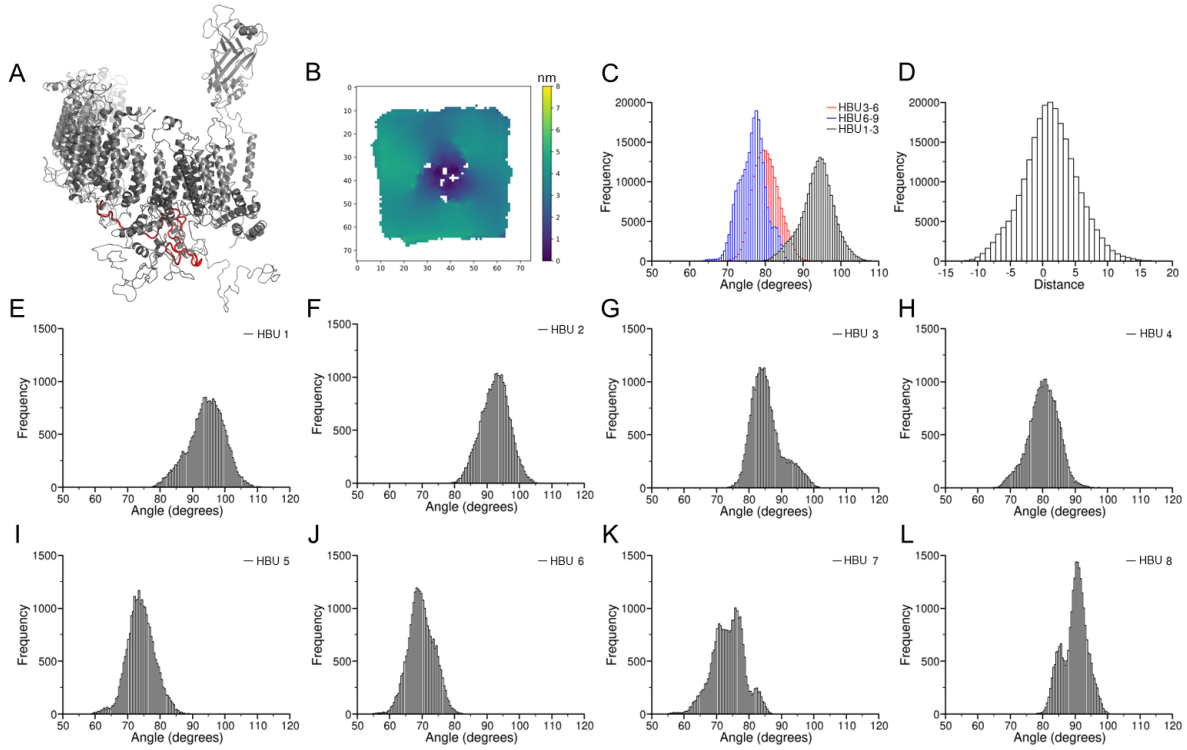


Figure 7.11 The Loss of the Beam Impacts the Z Angle of the Blade
A) The structure of a single subunit with the unstructured loop shown as a red tube. **B)** The heightmap of the mPIEZO1$_{Beam}$ system, **C)** the Z angles of the three sections of the blade and **D)** the distance between the centre of geometry of HBU1 and HBU4 in the Z direction. **E-L)** The Z angle of each of the HBUs in the mPIEZO1$_{Beam}$ system.

in which the beam is truncated with HBU7, so perhaps the beam was causing curvature of the membrane around the proximal region of the protein. The Z angle of the loop in the mPIEZO1$_{Beam}$ system was measured, demonstrating increased flexibility, but also that the angle still occupied 3° to 90°, with frequency occupied angles at 36°, 50° and 76°, which should have allowed some level of bilayer curvature.

As the angle of the beam could not be linked to the change in depth investigated earlier in this chapter, the Z angle of HBU1-3, 3-6 and 6-9 were measured as previously. The angle of the proximal three HBUs showed a shift towards wider angles, as associated with a flattening of the protein structure, as seen with the truncated protein structures. Interestingly, the range of angles these three HBUs can adopt increased upon loss of the structured beam beneath them, but not as much as the range of angles the loop itself could explore. Whereas the angle of the HBUs 3 to 6 were not substantially changed compared to the mPIEZO1$_{full\text{-}length}$ protein, maintaining their main angle of 80° (ranging between 70° and 91°). Despite this, changes occurred at the distal end of the blade, as HBUs 1 to 3 also shifted to wider angles of primarily 95° (ranging from 80° to 107°), rather than a main angle of 89° in the mPIEZO1$_{full\text{-}length}$ simulations. However, this did not appear to correspond to an increased downward projection of the distal 3 HBUs. Instead, the distance of the centre of geometry of HBU 1 and the centre of geometry of HBU4 in the Z direction was between -12 Å and 16 Å, but mainly occupied a

position of 1 Å in an upward direction. This could have been due to the lack of depth not requiring this region of the protein to occupy such a downward slope in comparison to the mPIEZO1$_{full\text{-}length}$ system.

To understand how the loss of beam structure had caused this effect on the angle of the HBUs, as there did not appear to be a large change in the flexibility of the three-HBU segments, the Z angle of each of the HBUs was measured, as in the mPIEZO1$_{full\text{-}length}$ system previously. Predictably, the most dramatic changes were in close proximity to the beam; HBU7 and HBU8 were both more flexible in the Z plane, as well as shifting to wider angles, causing a flattening of the blade in this region. HBU7 adopted angles up to 87°, whereas HBU8 was primarily at 90° compared to HBU9 at 82°, but was able to adopt angles up to 100°, consistent with membrane flattening and a downward bend in the blade. HBU6 also shifted to a larger angle of primarily 68°, but was less impacted by the loss of the structured beam region, likely due to not being in direct contact with this MPH. This change in Z angle of HBU6 and HBU7 were responsible for the loss of depth and confirmed their role in the creation of the membrane dome around mPIEZO1, but also demonstrated the role of the beam region in maintaining the upward direction of HBU6 and HBU7.

Further from the beam, at the distal region of the blade, there are smaller changes to the angle of HBU1, 2, 3 and 4, all of which adopt slightly wider Z angle conformations compared to the mPIEZO1$_{full\text{-}length}$ system, indicating that the changes at the proximal end of the blade are conferred to the distal end. How these changes are propagated is unclear, but this overall change in Z angle behaviour between the HBUs may be due to loss of depth causing the blade to have to adopt a flatter configuration and reducing the need or ability for the distal helical bundles to modulate the depth. This reveals a role for the beam in supporting the HBUs 6 to 9 and maintaining their angle against the Z axis and thereby assisting the formation of the dome.

7.2.6.2 The Role of the Beam in Protein Dynamics

The dynamics of the protein was studied as before, using RMSD/C_α and RMSF/C_α measures, which demonstrated the increased flexibility of the beam region and the overall RMSD/C_α (**Fig. 7.12**). The 2D protein density was of particular interest as there were regions of increased density at the point where the blade curves. Although this was present in other systems, the increased presence of this was of particular interest, as this region is HBU6 and was a point of flexibility in the mPIEZO1$_{full\text{-}length}$ system, so potentially reduced flexibility in the XY plane was of interest. Therefore, I measured the three-point XY angle of each of the HBUs, as previously, to assess how their flexibility had changed upon loss of the structured beam region.

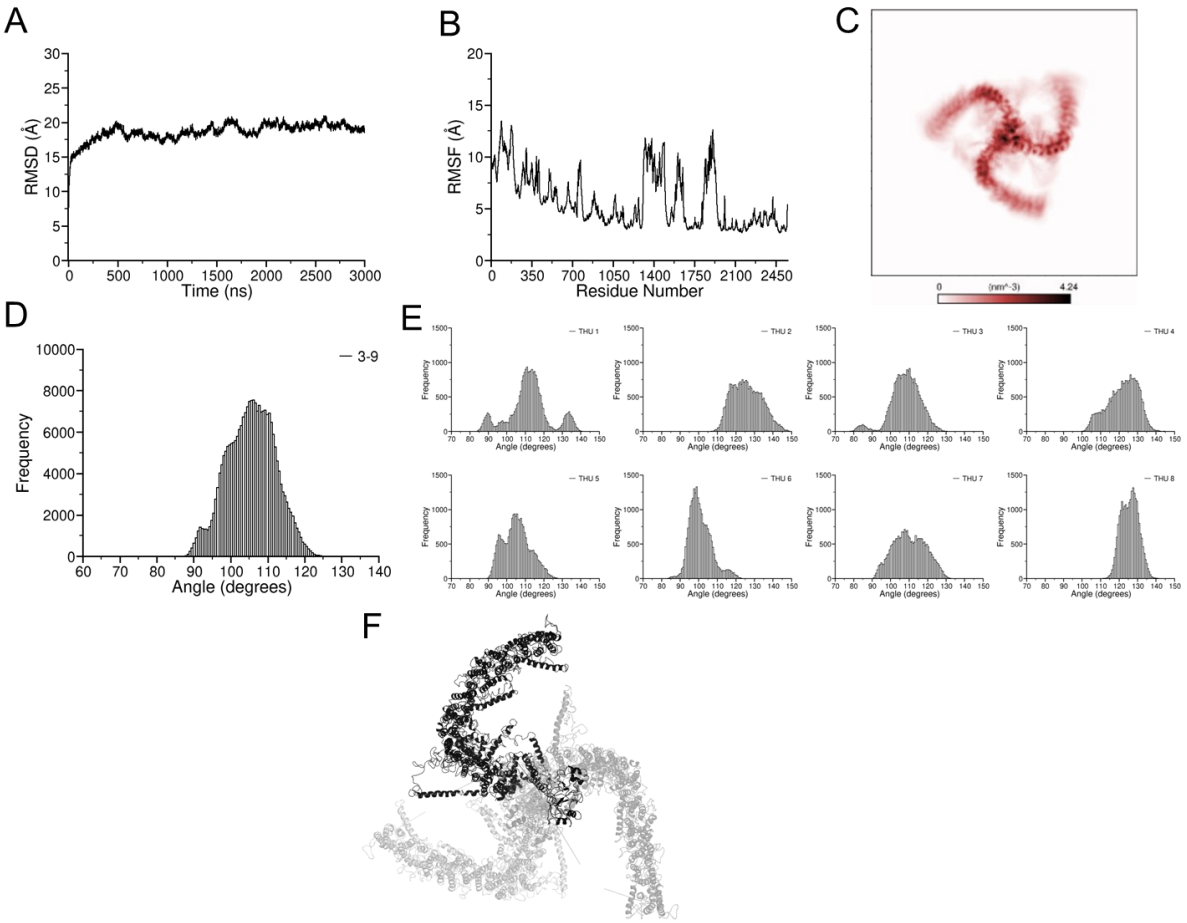


Figure 7.12 The Loss of the Beam Impacts the XY Angle of the Blade
The **A)** RMSD/C_α and **B)** RMSF/C_α of the backbone atoms of the mPIEZO1$_{Beam}$ protein averaged between replicates, and subunits in the case of the RMSF/C_α. **C)** The 2D density of the backbone particles of the mPIEZO1$_{Beam}$ protein. **D)** The XY angle between HBU3-6-9 and **E)** the XY angle of each of the HBUs. **F)** The structure of the blade when chain A is at 106, where the other subunits are shown as transparent QuikSurf.

The results were very striking, as all of the HBUs were more flexible in comparison to the mPIEZO1$_{full\text{-}length}$ system, with the largest increase for HBU1, which was able to range from 84° to 140° in the XY plane, with peaks at 89°, 111° and 134°. Many of the HBUs lost their major populated angle conformation, such as HBU2, HBU4 and HBU7. Comparatively, HBU6 did not have such an increased flexibility in the XY plane, spanning from only 84° to 123°, but primarily occupying an angle of 99°. This reduced flexibility in comparison to the rest of the HBUs is likely what causes it to have higher density in the 2D density plot.

Due to this change in XY flexibility of the HBUs, the overall angle of the blade was also investigated to assess how this was altered. Despite the increased flexibility of the individual HBUs, the overall blade angle was not more varied, rather it was only able to occupy one major state at 106°. Comparison of the blade structure between the mPIEZO1$_{Beam}$ and mPIEZO1$_{full\text{-}length}$ systems demonstrated that the initial three HBUs were somewhat contracted towards the channel pore, and the remainder of the blade continued to curve around. However, it was unable to adopt wider angles due to the increased rigidity of HBU6, as it

populates a more compact 99° angle in the mPIEZO1$_{Beam}$ system compared to the mPIEZO1$_{full-lengh}$ protein, in which the HBU angle ranges between 122° and 170°, with populated states at 130° and 157°. Additionally, HBU4 was able to occupy much wider angles, between 101° and 144° compared to 151° to 180° in the mPIEZO1$_{full-length}$ simulations, which explains the apparent straightening of the distal HBUs.

To investigate whether change in arrangement and dynamics of the blade upon loss of a structured beam affected the interactions between the MPHs of each mPIEZO1 chain, the number of contacts between each of the MPHs was measured over time. Despite the overall protein structure becoming more compact, potentially moving the MPHs into closer proximity for interactions, the number of contacts between the MPH from HBU2 and the posterior MPH from HBU8 was slightly decreased to 54.4% of the simulation time. However, changes were seen in the interactions between the posterior and anterior MPH from HBU8, as there were increased contacts between these regions of the protein in the mPIEZO1$_{Beam}$ system compared to the mPIEZO1$_{full-length}$ simulations, from 0.6% to 2.5% of the simulation time.

7.2.6.3 Lipid Interactions

Replacing the beam with an unstructured loop was not predicted to result in more interactions between the protein and lipids, but the lipid contact analysis revealed that the lipid contacts increased at all areas of the protein that are in contact with the membrane (**Fig. 7.13**). These were seen primarily in the PIP_2 interaction profile, but also with cholesterol. This overall increase appears to be due to the increase in flexibility of each of the HBUs, as lipids are able to interact at the interfaces of HBUs (Yang *et al*., 2022). Therefore, this increased motion of each HBU relative to the next may reveal an increased surface for interactions compared to the mPIEZO1$_{full-length}$ system, and interaction sites that are otherwise obscured by the presence of neighbouring protein. As the mPIEZO1$_{Beam}$ system is flatter than the mPIEZO1$_{full-length}$

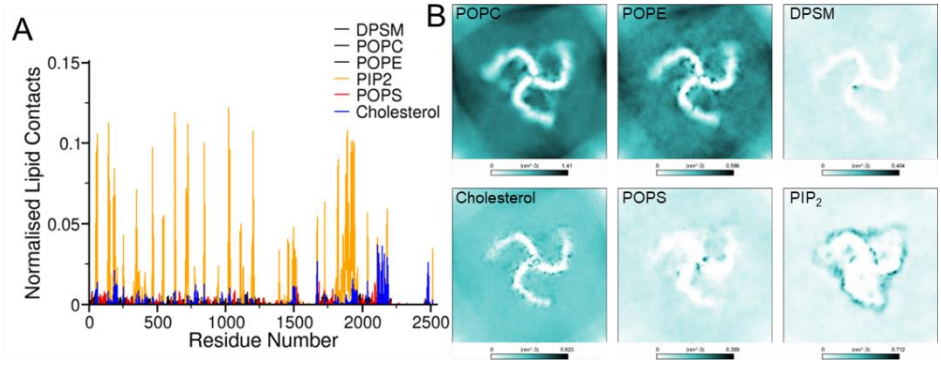


Figure 7.13 The Lipid Interactions with the mPIEZO1$_{Beam}$ Protein
A) The contacts between lipids and the protein normalised by the number of frames and lipid in the bilayer and averaged between replicates and chains, and **B)** the 2D density of the PO4 atom of the lipids in the bilayer mPIEZO1$_{Beam}$ system.

system, the lack of steep curvature may also have encouraged the diffusion of lipids around the protein and therefore led to more frequent contacts with different lipids.

7.2.7 Atomistic Investigation of the Activation of Truncated mPIEZO1

To understand the role of the factors identified previously in this chapter, such as blade compaction, increase and decrease in dome depth and flexibility of the HBUs, on mPIEZO1 channel activation in response to mechanical stimuli, atomistic simulations, as performed by De Vecchis and colleagues (2021), were performed. Here, final representative frames of the systems were converted to atomistic representations using the backward protocol (Wassenaar *et al*., 2014). The resulting systems were prepared for simulation by increasing their box size in the Z direction by up to 50%, resolvating the system and adding 50 calcium ions. Following a multi-step equilibration, the systems were equilibrated for 50 ns under constant pressure in

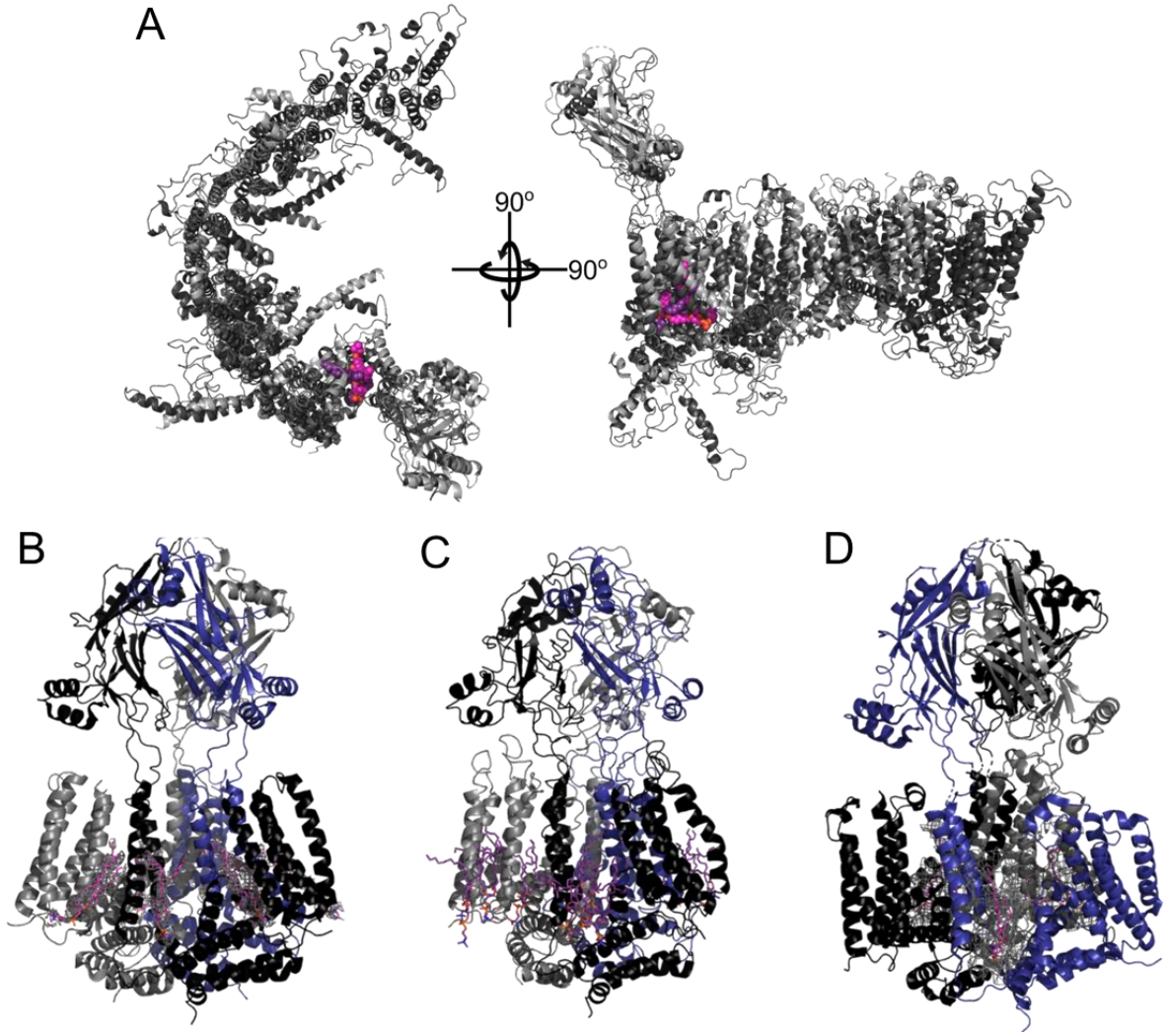


Figure 7.145Comparison Between the mPIEZO1 Atomistic Systems and Published Structures
A) A single subunit of the flattened mPIEZO1$_{full\text{-}length}$ system following 50 ns of simulation with -20 bar pressure on the membrane in dark grey with the published flattened mPIEZO1 structure (PDB: 7WLU) in light grey. The lipids in the pore are represented in spheres where the lipid from the simulation is in dark purple and the lipid from the structure is in magenta. The arrangement of the lipids in the pore in the **B)** flattened mPIEZO1 structure (PDB: 7WLU), **C)** the mPIEZO1$_{full\text{-}length}$ protein following 3 µs CG simulation and conversion to atomistic representation, and **D)** the curved mPIEZO1 structure (PDB: 7WLT).

the XY plane of the membrane to mimic mechanical force on the membrane. This pressure ranged from 1 to -40 bar. Previously this has been used to mimic mechanical force on the membrane and trigger mPIEZO1 opening, which is exemplified by flattening of the membrane dome and blades, twisting of the cap domain and rearrangement of the pore lining helices, thereby allowing solvent entrance to the pore (De Vecchis *et al*., 2021).

In this work, to understand whether these processes are altered in the truncated systems, the mPIEZO1$_{\text{full-length}}$, mPIEZO1$_{\Delta\text{HBU1-3}}$ and mPIEZO1$_{\text{Beam}}$ systems were backmapped to compare their activation. These were carried out at 0, -20 and -35 bar initially and although the membrane and structures were seen to flatten, there was no solvation of the pore or alteration of the structure of the pore lining helices, as was seen in previous simulations (De Vecchis *et al*., 2021). To assess whether this could be rectified by increasing the force on the membrane, simulations were run at -40 bar, but this resulted in membrane deformation and rupture within the first 10 ns. Unfortunately, this meant that I was unable to study the effect of truncation and loss of beam structure upon mPIEZO1 channel opening and activation by lateral membrane tension.

Despite this, the flattened mPIEZO1$_{\text{full-length}}$ structure did resemble the recently published cryo-EM resolved flattened structure (PDB: 7WLU) (Yang *et al*., 2022) (**Fig. 7.14**), as it has lipids in the pore and a similar arrangement of helices**,** resulting in a RMSD of only 6.3 Å, which was primarily due to a tilt in the structure of the extracellular cap domain. However, due to the low resolution of the flat mPIEZO1 structure, it was difficult to draw any concrete conclusions from this. However, the evidence of flattening and structural similarities was very encouraging and perhaps longer simulation time or increased tension could have triggered opening of the mPIEZO1$_{\text{full-length}}$ structure and allowed investigation of the impact of truncation and resulting increases or decreases in depth or blade angle on the function of mPIEZO1 in simulation.

7.3 Discussion

7.3.1 Comparison of the Simulated mPIEZO1$_{\text{full-length}}$ Structure to Published Structures

These simulations of the mPIEZO1$_{\text{full-length}}$, truncated-blade versions of the structure and where the beam was replaced with a loop have provided a number of insights into mPIEZO1 structure and the function of the blade region in modulating the dome depth and dynamics of the protein. Comparison of the simulated model structure to cryo-EM derived structures demonstrates how MD simulation is able to provide these insights and improve upon our understanding when other techniques do not provide all of the information. Simulation allowed the mPIEZO1$_{\text{full-length}}$ protein to refine its organisation, and resulted in a model that differs significantly from the original model generated by Chong and colleagues (2021), as well as the model generated by

AlphaFold2 (Jumper *et al*., 2021; Varadi *et al*., 2022) and the structure of mPIEZO2 (Wang *et al*., 2019).

The change in arrangement of the distal three HBUs was particularly striking as other indications of the arrangement of these HBUs, such as the mPIEZO2 cryo-EM structure (PDB: 6KG7) (Wang *et al*., 2019) and the AlphaFold2 mPIEZO1 model (Varadi *et al*., 2022), do not demonstrate this downward tilt of the bundles, rather they follow the trajectory of the other blade helices upwards. This is most pronounced in the mPIEZO2 cryo-EM structure, but this may be explained by the generation of the cryo-EM map, which involved generating three focused electron density maps, one of which was of the distal helical bundles, which were then fitted together (Wang *et al*., 2019). Alternatively, the environment of the protein during structure determination may have allowed or encouraged the adoption of otherwise non-native blade conformations, as the protein was solubilised in the detergent glycoldiosgenin. Perhaps the lack of lateral pressure from the membrane and the effect of specific lipids may have altered the curvature of the blade, as detergent solubilisation is known to impact the structure of membrane proteins (Bill *et al*., 2011).

Additionally, the changes in MPH arrangement were of particular interest as they were a primary aspect of making sense of the data in this chapter. Several studies of the structure of mPIEZO1 and mPIEZO2 have identified a supporting and stabilising role for the beam, as it sits beneath the helical bundles 7, 8 and 9, and suggest the lack of a similar structural feature in the distal blade may be a reason for their increased flexibility (Wang *et al*., 2019). However, the shift in arrangement of the MPHs upon simulation may provide this stabilising and modulating effect. For instance, the MPHs from HBU 1, 3 and 4 move to sit beneath the HBUs and may directly provide stability to the HBU and maintain its position relative to the other HBUs in the blade. Additionally, the change in angle of these helices in the Z axis, may explain the downward slope of these distal HBUs, as their supporting scaffold is now also slanted in that direction in the Z axis.

7.3.2 Structural Mechanism for Modulation of mPIEZO1 dome depth

The data from this chapter indicates an essential role for the HBUs in modulating the depth of the dome formed by mPIEZO1 and the dynamics of the protein blade in the XY plane. Previous work comparing the full-length structure and a truncated mPIEZO1 has shown that the distal N-terminal bundles assist in formation of a deeper membrane dome (~6.5 nm) compared to the $mPIEZO1_{trunc}$ structure (~5.5 nm) (Chong *et al*., 2021) (note the discrepancy in measurement between previous work and this work is due to the calculation of the depths, using the same method the $mPIEZO1_{full\text{-}length}$ system in this chapter generates a dome of 6.41 nm), which corresponds to experimental dome depth measurements (~6 nm) (Guo and

MacKinnon, 2017). However, my work demonstrates the opposite effect; The loss of the first three HBU results in a deeper dome (8.22 nm using the comparable calculation to the previously published work), rather than a shallower one, due to the downward slant of the first three HBUs.

Initially, this result was puzzling, as a comparable number of residues were removed from my mPIEZO1$_{\Delta HBU1\text{-}3}$ structure (550 residues) as are unresolved in the cryo-EM published structure (576 residues), and the 26 additional residues are an unstructured loop region into the first helix of HBU4, which was seen to form a long helix that sits under the corresponding HBU in the simulations, and so was unlikely to be altering the membrane dome formation. Therefore, attention was turned to other regions of the mPIEZO1 structure that were missing in the MD investigation of the cryo-EM truncated protein (Chong *et al*., 2021), these included the MPHs from HBU5, HBU7 and the posterior MPH from HBU8, as well as disordered intracellular loops.

The absence of these MPH was of particular interest, as their role in modulating the flexibility and angle of the blade became apparent during these simulations. Therefore, perhaps their absence in the previously published mPIEZO1$_{trunc}$ vs mPIEZO1$_{full\text{-}length}$ studies was the main driver for the changes in depth that were seen. The MPH absence may have the same impact on the mPIEZO1$_{trunc}$ systems, as the loss of the structured beam region in the mPIEZO1$_{Beam}$ simulations. This may have resulted in lack of stability of the HBUs and disruption of their angle against the Z axis and may have resulted in loss of dome depth compared to the full-length model, which was attributed to the absence of the distal three HBUs, rather than the MPHs. This hypothesis is strengthened by the evidence that amphipathic helices such as these are thought to be able to introduce curvature to membranes (Drin and Antonny, 2010; Saotome *et al*., 2018), and are a key mechanism for bilayer tension-mediated gating in the MscL mechanosensitive ion channel family (Bavi *et al*., 2016). In addition to this, these are sites of extensive interaction with PIP_2, which is functionally important for mPIEZO1 activity (Borbiro *et al*., 2015). The extensive PIP_2 interactions with the MPHs, especially the posterior MPH from HBU8, which was modulated by blade length and dome depth, may explain the difference in results between the mPIEZO1$_{trunc}$ and mPIEZO1$_{\Delta HBU1\text{-}3}$ structures, as PIP_2 is an inverse conical lipid (Pomorski *et al*., 2014) and so could act to increase the curvature of the dome and resulting depth.

Despite the differences with other MD simulation studies comparing truncated and full-length mPIEZO1 studies, the work presented in this chapter presents a convincing bank of evidence for how the blade modulates the depth of the dome around mPIEZO1. Three regions of the blade modulate the depth separately; firstly, the HBUs 6-9 provide the main upward angle and contribute to the majority of the dome depth. Following these proximal HBUs, the HBUs 3-6

provide some upward slope to the blade, but their role is mainly to compensate for changes in the HBU 1-3 angle, which is primarily downward in the membrane, thereby modulating the overall depth of the dome and causing the bilayer beyond the protein to continue to be perturbed by the blade. Each of the angles against the Z axis of these regions appears to be modulated by MPHs, either the MPH from HBU1 and 3, which slope downwards below these HBUs and allow them to adopt their downward configuration, or the MPH from HBU5 which controls the angle of HBU 3, 4 and 5, or the beam, which maintains the upward direction of HBU6 and HBU7.

7.3.3 Structural Mechanism for the Modulation of mPIEZO1 Blade Dynamics

The dynamics of the mPIEZO1 blade in the XY plane has only been investigated in simulations of the atomistic protein (De Vecchis *et al*., 2021), in which their straightening was identified. However, the data in this chapter has identified 3 main blade conformations in the $mPIEZO1_{full\text{-}length}$ structure, which are reduced to a single main conformation upon loss of the first two HBUs. Identification of the points of flexibility in the XY plane along the blade made clear what was driving the mPIEZO1 dynamics. The MPH along the blade appeared to be modulating the angle each of the HBUs could adopt in the XY plane, the beam prevented the lateral movement of HBU6, 7, 8 and 9, and upon loss of this MPH these HBUs caused the blade to become more compact, while increasing the flexibility of each HBU. The MPH from HBU5 maintained the inward curve of the HBUs 3 to 5 and prevented over flexion of the blade in either direction by forming protein contacts with the HBUs alongside it.

The most distal anterior facing MPH, the MPH from HBU2, appeared to play a role in lipid-mediated inter-subunit interactions, along with the anterior and posterior MPH from HBU8, and on occasion the MPH from HBU5. This again demonstrates the role of the MPHs in maintaining blade structure, as these interactions are associated with different blade conformations seen throughout the simulations, and the points of flexibility along the blade lack direct contact with a stabilising HBU. This finding appears to be consistent with experimental investigation of the mPIEZO1 channel, as Lin and colleagues (2019) saw that the blades occupied an angle of ~120° in POPE and POPG bilayers, and the structural evidence of the blade forming distinct conformations in the XY plane (Saotome *et al*., 2018), thereby confirming that this is an angle the blade is able to occupy and potentially the need for PIP_2 to form these lipid-mediated interactions.

7.3.4 Lipid Interactions with Truncated mPIEZO1

The lipid interactions in this work did not reveal heretofore unidentified lipid binding sites along the blade, as was anticipated at the beginning of this work, rather the changes in lipid binding appeared to be a consequence of the changes in protein dynamics and membrane curvature.

This was primarily seen in the lipid interactions with the pore of the protein, as the blade length decreased, interactions between anionic lipids and the pore increased, likely due to increased accessibility of this region of the protein. The truncation of the blade caused the opening to the anterior side of the blade to increase and diffusion of lipids into this space is easier. Additionally, the reduced curvature around the protein itself likely improved this, as this curvature represents an energy barrier for diffusion (Woodward *et al*., 2018). Due to the majority of the lipid interactions seeming to be a result of other changes to the protein, few conclusions can be drawn about their role in the structure and function of the blade, but the concurrence with the lipids resolved in mPIEZO1 cryo-EM studies (Yang *et al*., 2022), indicates that the effects seen here are reliable.

7.3.5 Biological Relevance of the mPIEZO1 Blade Dynamics and Structure

The blade is considered to be the region of PIEZO channels which is directly responsible for sensing and triggering the response to mechanical stimuli (Wu *et al*., 2016; Zhao *et al*., 2018; Wang *et al*., 2018; Lin *et al*., 2019; Yang *et al*., 2022), and so understanding the motions and conformations of the blade is essential to understanding how the protein might be activated. The presence of multiple points of flexibility along the blade in both the Z and XY directions begins to demonstrate the complexity of mPIEZO1 activation, as each of the blade regions may respond to mechanical force differently. Thereby, the different regions of the blade itself may act as gates and only allow force propagation to the pore if it reaches a certain threshold, which may be modulated by the conformation of the blade in different biological circumstances. This would be essential to controlling the function of mPIEZO1, as certain cells will be constantly under some force from their surroundings, and so dictating when a mechanical stimulus is sufficient for activation is crucial for preventing overactivity of the channel.

Additionally, the direction of the distal three HBUs in the bilayer and their ability to cause perturbation of the membrane beyond the channel itself may be important for clustering of PIEZO1 channels, as has been seen in super-resolution microscopy studies (Ridone *et al*., 2020). The continuation of the membrane curvature into the surrounding bilayer may allow the PIEZO1 proteins to interact through regions of increased or decreased curvature and form larger scale organisation of the channels and bilayer. As cholesterol is able to disrupt this (Ridone *et al*., 2020), and concentration of this lipid has been identified as a driver of dome depth in simulation (Chong *et al*., 2021), the ability of the blade to alter the extension of the membrane deformation may provide the protein another mechanism to modulate its function.

Finally, the dynamics of the blade in the XY direction may be a mechanism for controlling activation, as the straightening of the blades is seen upon flattening and pore solvation (De

Vecchis *et al*., 2021; Yang *et al*., 2022). Therefore, being at a wider angle may make the channel more sensitive to stimuli, or less so at narrower angles. This would also provide the cell with a mechanism by which to modulate the sensitivity of the PIEZO1 channels, as the motion of the blade in the XY direction appears to be linked to PIP_2 interactions, which have been implicated in the function of mPIEZO1 (Borbiro *et al*., 2015). Controlling the local PIP_2 concentration, especially at the MPH and pore, may allow the blade to adopt more acute or obtuse angles, as needed by the cell, or allow differences in PIEZO1 activity between cell type and lipid composition.

7.3.6 Future Directions

This work has led to a number of hypotheses surrounding the role of both the HBUs and the MPHs in blade function and structure. However, testing of these hypotheses is essential for fine-tuning further mPIEZO1 structural and functional study. Firstly, clarifying the role of the other MPHs identified in this work in blade dynamics in the Z and XY planes should be performed. Replacing the beam with an unstructured loop was key to understanding its role in dome depth and stabilising the proximal HBUs, and so replacing the MPH from HBU2 and the MPH from HBU5 with unstructured loops should also result in confirmation of their role in MD simulation. Additionally, applying restraints to the points of flexibility along the mPIEZO1 blade may allow better delineation of the contribution of each of these to mPIEZO1 blade dynamics, as often in these simulations the motions appeared to be cooperative and simultaneous. Another area of simulation that should be further pursued is the atomistic lateral tension simulations and their ability to demonstrate changes in the activation of truncated or mutated mPIEZO1.

Experimental validation of these findings could be achieved in functional studies of mPIEZO1 through mutation of the interfaces between HBUs to introduce disulphide bridges or other cross-links to reduce their relative motion. The results of this could be studied functionally though Ca^{2+}-imaging assays and electrophysiology, whereas structural changes could be assessed through HS-AFM as by Lin and colleagues (2019), who were able to identify the 120° periodicity of the mPIEZO1 blades. Alternatively, if disulphide bridges between subunits was able to trap blade conformations, this may facilitate structurally resolution of the distal HBUs through cryo-EM, which in combination with liposomes and native lipids could allow further structural study of the channel.

7.4 Conclusions

To understand the structure and dynamics of the full-length mPIEZO1 blade, a rational approach was used to truncate the blade with the aim of identifying roles for each of the HBUs. Truncating the blade had significant effects on the depth profile of the membrane dome around

mPIEZO1, which followed a non-linear relationship to blade length, and demonstrated a role for each of the HBUs in modulating the formation of the dome. The dynamics of the blade in the XY plane demonstrated the ability of the blade to adopt three main conformations, but loss of the HBUs abolishes ability of the blade to do this and led to one main conformation. Looking at the changes in Z and XY angle on a HBU-by-HBU basis led to the identification of points of flexibility at HBU1, HBU3, HBU4 and HBU6 and the indication that the MPH play a significant role in controlling the stability and angle of the HBUs they are associated with. Overall, the data presented in this chapter makes clear that the structure of the blade is more complex than a single vector from the channel pore, and so the subtleties of its arrangement should not be overlooked, but rather studied closer to understand how it can impact the protein function on a molecular level.

8. Summary and Outlook

In this thesis, proteins from the mPPase family of membrane-bound enzymes and from the PIEZO mechanosensitive ion channel family were selected and examined through multiscale simulations to understand how their unique features and interactions with their environment contribute to their function and structure. Despite appearing to be disparate protein families, expressed in different organisms, performing different cellular functions following different activation mechanisms, there are marked similarities between the two. Both are involved in cation translocation, exist as homoligomeric assemblies, are linked to changes in osmotic pressure and are clinically relevant. This allows similar questions to be explored in different settings and provide wider insights into membrane protein function than simply focusing on a single biomolecule, membrane or organism.

Firstly, I focused on the protein-lipid interactions between different members of the mPPase family. This resulted in evidence that mPPases possess a preference for anionic lipids and that these interactions occur at specific sites on the surface of the protein in contact with the cytoplasmic leaflet of the membrane. These were attributed to the positive inside rule for alpha-helical proteins, but the presence of specific binding sites and preference for particular lipid species, such as POPA, initiated the investigation of lipid-interaction sites in other structurally resolved members of the mPPase family from evolutionarily diverse organisms. This led to the prediction that binding sites for anionic lipids at the interface between the two subunits and the distal end of the proteins are conserved across species. These sites are made up of structurally comparable arrangements of residues on helices that are key for both functional and conformational aspects of the catalytic cycle, indicating their importance for mPPase activity. The interfacial site was hypothesised to be involved in inter-subunit communication and conformational changes, whereas the distal site may form stabilising interactions with the bilayer.

Then the dynamics of different states of the mPPase catalytic cycle were investigated to understand the changes to the protein as it moves through the different states. Ideally, conformational changes or evidence of asymmetry would have been observed in this work, but due to the lack of dynamics in the substrate-analogue bound structure, even upon removal of the contents of the active site, this was not realised in this work. However, further insights were gained into the closure of the 5-6 loop over the active site and how this is maintained in the substrate analogue-bound system compared to the resting structure. It is maintained not only contacts between the active site contents and the loop, but also the increase in charge of the active site causes the hydrophobic residues on the 5-6 loop to form a hydrophobic interface with the 13-14 loop, which sits over the 5-6 loop in simulation. In addition to the

insights into the substrate analogue-bound state of the catalytic cycle, a mode of sodium ion entry and binding to the ionic gate was defined using insights from the simulations of the resting state. Upon binding of the sodium ion, a series of conformational changes occur at the ionic gate and in the active site, which was exploited for *in silico* high-throughput screening of potential inhibitors. Screening of structures determined through simulation, such as the sodium-bound resting state and "empty" substrate analogue-bound state resulted in the identification of a novel inhibitor with the best IC_{50} to date. Alongside this, which may provide another tool to investigate mPPases, the catalytic cycle could be supplemented with key insights from this work.

Finally, a full-length model of mPIEZO1 was systematically truncated to understand the role of different regions of the blade in the structure and function of this mechanosensitive ion channel. Despite previous comparison of the full-length model used in this work to the comparatively truncated mPIEZO1 structure published by Guo and MacKinnon (2017), the study described in this thesis was able to provide more specific understanding of the role of each helical bundle unit (HBU) and how they impacted the motions of the blade in all XYZ planes. Comparison to the previous work resulted in the most surprising finding: the loss of the first three HBUs resulted in a deeper dome rather than a shallower one. Analysis of the dynamics of the blade demonstrated that this was due to there being several points of flexibility along the blade in both the Z and XY directions, determined by the proximity of a membrane parallel helix (MPH). Loss of these points of flexibility could not be fully compensated for by the other HBUs and resulted in "sinking" of the protein into the membrane. Alongside this, loss of the distal HBUs attenuated the dynamics of the blade in the XY plane, reducing the populated conformations from three to just one. The role of the MPHs in modulating the structure and angle of regions of the blade was somewhat confirmed by substitution of the structured beam helix with an unstructured loop, which resulted in loss of regulation of the blade dynamics. These observations could not be fully investigated in atomistic simulations, as the flattened structures did not appear to open as seen by De Vecchis and colleagues (2021). Despite this, the insights into blade structure and its modulation provide more depth to the understanding of the otherwise neglected region of the PIEZO1 protein.

The time restraints of the project, the complications due to the Covid-19 pandemic and changes to my supervision team and project, meant several interesting areas of work presented in this thesis could not be fully explored, but this opened up avenues for future work. For the mPPase research, the lipid interactions should be investigated *in vitro*, likely through an approach combining mutagenesis of the sites with relipidation following protein production and purification. Characterisation using mass spectrometry to confirm lipid binding, stability screening and functional screening can be used to assess the role lipids play in mPPase

activity. Additionally, if stabilisation is a role lipids play, screening of less stable homologues and subsequent structure determination attempts could be a way to diversify the structures of mPPases available. In addition to this mPPase investigation, there are a number of open questions following the data presented in this thesis. Firstly, further investigation and confirmation of the sodium binding mechanism should be performed, perhaps through mutagenesis of the $R^{5.50}$ residue to a negatively charged or neutral amino acid to examine the effects on sodium transport and perhaps also the kinetics of the enzyme, as this process may represent a bottle neck for catalytic "reset". Of even more interest are the inhibitors identified by screening simulated structures. The first step for further optimisation of the 7795842 compound should be determination of the structure of a 7795842-bound mPPase. This would allow better understanding of the binding site and mechanism and more rational inhibitor discovery approaches. Following this, identification of analogues and a structure-activity relationship library can be generated and examined.

The results presented in this work, in which I examined the role and dynamics of the HBUs of the mPIEZO1 blade also should be examined further. This chapter demonstrated a role for the MPHs, which was only somewhat examined by replacing the structured beam helix with a loop, but this should be performed for the other MPHs in simulations to see how this impacts the blade dynamics and structure. Also, the location of the MPHs, as they move in simulations from their modelled positions, should be determined experimentally. Perhaps through hydrogen-deuterium exchange mass spectrometry to examine the solvent accessible surface of the molecule. Also, to experimentally validate the dynamics of the blade, spin labelling of HBUs or MPHs could be used to determine the range of distances they can adopt. As in the mPPase study, the experimental and computational data could be compared to understand the distances on a molecular level and further validate the use of simulations to study PIEZO ion channels. Alternatively, using AFM to understand the mPIEZO1 dynamics and changes upon activation may be able to reveal more about the dynamics of mPIEZO1 in bilayers. This work may also lead to an experimentally determined, full-length PIEZO1 channel structure. As the main barriers to this appear to be the flexibility of the distal regions of the blades, perhaps introducing cross-links between the HBUs that form the points of flexibility would reduce this motion and trap the PIEZO1 blade for structure determination, although this might be at the cost of reduced function. The effects on mPIEZO1 activation can be confirmed through electrophysiology alongside these mutations, which would identify important resides along the blades for mPIEZO1 function. Finally, the effect of loss of the HBUs and the resulting changes in depth and dynamics should be investigated atomistically, perhaps through longer timescales, higher tensions or different starting configurations the changes on opening could be examined.

The data presented in this thesis deepen our understanding of two important biomolecules through simulation techniques. Although the insights gained through study of each protein is not directly translatable to the other, by examining two evolutionarily and mechanistically different systems, the knowledge gained here might be transferable to a much larger diversity of membrane proteins. Whether through the identification of amphipathic helices in controlling membrane curvature and modulation of the motion of domains, or the potential of lipids to modulate the function and stability of a large family of clinically relevant membrane proteins.

9. References

Almén, M.S., Nordström, K.J.V., Fredriksson, R. and Schiöth, H.B. 2009. Mapping the human membrane proteome: A majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biology*. **7**(1), p.50.

Anashkin, V.A., Malinen, A.M., Bogachev, A. V. and Baykov, A.A. 2021. Catalytic asymmetry in homodimeric H^+-pumping membrane pyrophosphatase demonstrated by non-hydrolyzable pyrophosphate analogs. *International Journal of Molecular Sciences*. **22**(18).

Andréll, J. and Tate, C.G. 2013. Overexpression of membrane proteins in mammalian cells for structural studies. *Molecular Membrane Biology*. **30**(1), pp.52–63.

Arkhipova, V., Guskov, A. and Slotboom, D.-J. 2017. Analysis of the Quality of Crystallographic Data and the Limitations of Structural Models. *Journal of General Physiology*. **149**(12), pp.1091–1103.

Arnarez, C., Mazat, J.-P., Elezgaray, J., Marrink, S.-J. and Periole, X. 2013. Evidence for Cardiolipin Binding Sites on the Membrane-Exposed Surface of the Cytochrome bc_1. *Journal of the American Chemical Society*. **135**(8), pp.3112–3120.

Artukka, E., Luoto, H.H., Baykov, A.A., Lahti, R. and Malinen, A.M. 2018. Role of the Potassium/Lysine Cationic Center in Catalysis and Functional Asymmetry in Membrane-Bound Pyrophosphatases. *Biochemical Journal*. **475**(6), pp.1141–1158.

Asaoka, M., Segami, S. and Maeshima, M. 2014. Identification of the Critical Residues for the Function of Vacuolar H^+-Pyrophosphatase by Mutational Analysis Based on the 3D Structure. *Journal of Biochemistry*. **156**(6), pp.333–344.

Asaoka, M.M.A., Segami, S., Ferjani, A. and Maeshima, M. 2016. Contribution of PPi-Hydrolyzing Function of Vacuolar H^+-Pyrophosphatase in Vegetative Growth of Arabidopsis: Evidenced by Expression of Uncoupling Mutated Enzymes. *Frontiers in Plant Science*. **7**(415), pp.1–12.

Au, K.M., Barabote, R.D., Hu, K.Y. and Jr, M.H.S. 2006. Evolutionary Appearance of H^+-Translocating Pyrophosphatases. *Microbiology*. **152**(5), pp.1243–1247.

Bae, C., Gnanasambandam, R., Nicolai, C., Sachs, F. and Gottlieb, P.A. 2013. Xerocytosis is caused by mutations that alter the kinetics of the mechanosensitive channel PIEZO1. *Proceedings of the National Academy of Sciences of the United States of America*. **110**(12).

Bae, C., Gottlieb, P.A. and Sachs, F. 2013. Human PIEZO1: Removing inactivation. *Biophysical Journal*. **105**(4), pp.880–886.

Bae, C., Sachs, F. and Gottlieb, P.A. 2015. Protonation of the human PIEZO1 ion channel stabilizes inactivation. *Journal of Biological Chemistry*. **290**(8), pp.5167–5173.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J. and McCammon, J.A. 2001. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America*. **98**(18), pp.10037–10041.

Ballesteros, J.A. and Weinstein, H. 1995. Integrated Methods for the Construction of Three-Dimensional Models and Computational Probing of Structure-Function Relations in G Protein-Coupled Receptors. *Methods in Neurosciences*. **25**(19), pp.366–428.

Baltscheffsky, M., Schultz, A. and Baltscheffsky, H. 1999. H^+-Proton-Pumping Inorganic Pyrophosphatase: A Tightly Membrane-Bound Family. *FEBS Letters*. **452**(3), pp.121–127.

Bartoli, F., Debant, M., Chuntharpursat-Bon, E., Evans, E.L., Musialowski, K.E., Parsonage, G., Morley, L.C., Simon Futers, T., Sukumar, P., Scott Bowen, T., Kearney, M.T., Lichtenstein, L., Roberts, L.D. and Beech, D.J. 2022. Endothelial Piezo1 sustains muscle capillary density and contributes to physical activity. *Journal of Clinical Investigation*. **132**(5).

Bauer, P., Hess, B. and Lindahl, E. 2022. GROMACS 2022.3 Manual.

Bavi, N., Cortes, D.M., Cox, C.D., Rohde, P.R., Liu, W., Deitmer, J.W., Bavi, O., Strop, P., Hill, A.P., Rees, D., Corry, B., Perozo, E. and Martinac, B. 2016. The role of MscL amphipathic N terminus indicates a blueprint for bilayer-mediated gating of mechanosensitive channels. *Nature Communications*. **7**(1), pp.1–13.

Baykov, A.A. 2020. Energy Coupling in Cation-Pumping Pyrophosphatase—Back to Mitchell. *Frontiers in Plant Science*. **11**(February), pp.1–5.

Baykov, A.A., Malinen, A.M., Luoto, H.H. and Lahti, R. 2013. Pyrophosphate-Fueled Na^+ and H^+ Transport in Prokaryotes. *Microbiology and Molecular Biology Reviews*. **77**(2), pp.267–276.

Belogurov, G.A. and Lahti, R. 2002. A Lysine Substitute for K^+: A460K Mutation Eliminates K^+ Dependence in H^+-Pyrophosphatase of *Carboxydothermus hydrogenoformans* . *Journal of Biological Chemistry*. **277**(51), pp.49651–49654.

Benkert, P., Biasini, M. and Schwede, T. 2011. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*. **27**(3), pp.343–350.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Research*. **28**(1), pp.235–242.

Best, R.B. 2019. Atomistic Force Fields for Proteins *In*: *Methods in Molecular Biology*. Humana Press Inc., pp.3–19.

Biggin, P.C. and Bond, P.J. 2008. Molecular dynamics simulations of membrane proteins. *Methods in Molecular Biology*. **443**, pp.147–160.

Bill, R.M., Henderson, P.J.F., Iwata, S., Kunji, E.R.S., Michel, H., Neutze, R., Newstead, S., Poolman, B., Tate, C.G. and Vogel, H. 2011. Overcoming Barriers to Membrane Protein Structure Determination. *Nature Biotechnology*. **29**(4), pp.335–340.

Borbiro, I., Badheka, D. and Rohacs, T. 2015. Activation of TRPV1 channels inhibits mechanosensitive piezo channel activity by depleting membrane phosphoinositides. *Science Signaling*. **8**(363).

Brini, F., Hanin, M., Mezghani, I., Berkowitz, G.A. and Masmoudi, K. 2006. Overexpression of Wheat Na^+/H^+ Antiporter TNHX1 and H^+-Pyrophosphatase TVP1 Improve Salt- and Drought-Stress Tolerance in *Arabidopsis thaliana* plants. *Journal of Experimental Botany*. **58**(2), pp.301–308.

Broecker, J., Eger, B.T. and Ernst, O.P. 2017. Crystallogenesis of Membrane Proteins Mediated by Polymer-Bounded Lipid Nanodiscs. *Structure*. **25**(2), pp.384–392.

Brooks, B.R., Brooks, C.L., Mackerell, A.D., Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A.R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R.W., Post, C.B., Pu, J.Z., Schaefer, M., Tidor, B., Venable, R.M., Woodcock, H.L., Wu, X., Yang, W., York, D.M. and Karplus, M. 2009. CHARMM: The Biomolecular Simulation Program. *Journal of Computational Chemistry*. **30**(10), pp.1545–1614.

Bruininks, B.M.H., Souza, P.C.T. and Marrink, S.J. 2019. *A Practical View of the Martini Force Field*.

Buchan, D.W.A. and Jones, D.T. 2019. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Research*. **47**(W1), pp.W402–W407.

Büscher, P., Cecchi, G., Jamonneau, V. and Priotto, G. 2017. Human African Trypanosomiasis. *The Lancet*. **390**(10110), pp.2397–2409.

Cahalan, S.M., Lukacs, V., Ranade, S.S., Chien, S., Bandell, M. and Patapoutian, A. 2015. Piezo1 links mechanical forces to red blood cell volume. *eLife*. **4**(MAY).

Cecchetti, C., Strauss, J., Stohrer, C., Naylor, C., Pryor, E., Hobbs, J., Tanley, S., Goldman, A. and Byrne, B. 2021. A novel high-throughput screen for identifying lipids that stabilise membrane proteins in detergent based solution. *PLoS ONE*. **16**(7 July).

Chong, J., De Vecchis, D., Hyman, A.J., Povstyan, O. V., Ludlow, M.J., Shi, J., Beech, D.J. and Kalli, A.C. 2021. Modeling of full-length Piezo1 suggests importance of the proximal N-terminus for dome structure. *Biophysical Journal*. **120**(8), pp.1343–1356.

Chubinskiy-Nadezhdin, V.I., Negulyaev, Y.A. and Morachevskaya, E.A. 2011. Cholesterol depletion-induced inhibition of stretch-activated channels is mediated via actin rearrangement. *Biochemical and Biophysical Research Communications*. **412**(1), pp.80–85.

Conrard, L., Stommen, A., Cloos, A.S., Steinkühler, J., Dimova, R., Pollet, H. and Tyteca, D. 2018. Spatial relationship and functional relevance of three lipid domain populations at the erythrocyte surface. *Cellular Physiology and Biochemistry*. **51**(4), pp.1544–1565.

Corey, D.P. and Hudspeth, A.J. 1979. Response latency of vertebrate hair cells. *Biophysical Journal*. **26**(3), pp.499–506.

Corey, R.A., Pyle, E., Allen, W.J., Watkins, D.W., Casiraghi, M., Miroux, B., Arechaga, I., Politis, A. and Collinson, I. 2018. Specific Cardiolipin–SecY Interactions are Required for Proton-Motive Force Stimulation of Protein Secretion. *Proceedings of the National Academy of Sciences*. **115**(31), pp.7967–7972.

Corradi, V., Mendez-Villuendas, E., Ingólfsson, H.I., Gu, R.X., Siuda, I., Melo, M.N., Moussatova, A., Degagné, L.J., Sejdiu, B.I., Singh, G., Wassenaar, T.A., Delgado Magnero, K., Marrink, S.J. and Tieleman, D.P. 2018. Lipid-Protein Interactions Are Unique Fingerprints for Membrane Proteins. *ACS Central Science*. **4**(6), pp.709–717.

Corradi, V., Sejdiu, B.I., Mesa-Galloso, H., Abdizadeh, H., Noskov, S.Y., Marrink, S.J. and Tieleman, D.P. 2019. Emerging Diversity in Lipid-Protein Interactions. *Chemical Reviews*.

Coste, B., Mathur, J., Schmidt, M., Earley, T.J., Ranade, S., Petrus, M.J., Dubin, A.E. and Patapoutian, A. 2010. Piezo1 and Piezo2 are essential components of distinct

mechanically activated cation channels. *Science*. **330**(6000), pp.55–60.

Cox, C.D., Bae, C., Ziegler, L., Hartley, S., Nikolova-Krstevski, V., Rohde, P.R., Ng, C.A., Sachs, F., Gottlieb, P.A. and Martinac, B. 2016. Removal of the mechanoprotective influence of the cytoskeleton reveals PIEZO1 is gated by bilayer tension. *Nature Communications*. **7**(1), pp.1–13.

Crompton, P.D., Moebius, J., Portugal, S., Waisberg, M., Hart, G., Garver, L.S., Miller, L.H., Barillas-Mury, C. and Pierce, S.K. 2014. Malaria Immunity in Man and Mosquito: Insights into Unsolved Mysteries of a Deadly Infectious Disease. *Annual Review of Immunology*. **32**(1), pp.157–187.

Dalghi, M.G., Ruiz, W.G., Clayton, D.R., Montalbetti, N., Daugherty, S.L., Beckel, J.M., Carattino, M.D. and Apodaca, G. 2021. Functional roles for PIEZO1 and PIEZO2 in urothelial mechanotransduction and lower urinary tract interoception. *JCI Insight*. **6**(19).

DeLano, W. 2002. Pymol: An Open-Source Molecular Graphics Tool. *CCP4 Newsletter On Protein Crystallography*. **40**.

Delmas, P., Parpaite, T. and Coste, B. 2022. PIEZO channels and newcomers in the mammalian mechanosensitive ion channel family. *Neuron*. **110**(17), pp.2713–2727.

Docampo, R., Jimenez, V., Lander, N., Li, Z.H. and Niyogi, S. 2013. New Insights into Roles of Acidocalcisomes and Contractile Vacuole Complex in Osmoregulation in Protists. *International Review of Cell and Molecular Biology*. **305**, pp.69–113.

Dolinsky, T.J., Nielsen, J.E., McCammon, J.A. and Baker, N.A. 2004. PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Research*. **32**(WEB SERVER ISS.), pp.W665–W667.

Drin, G. and Antonny, B. 2010. Amphipathic helices and membrane curvature. *FEBS Letters*. **584**(9), pp.1840–1847.

Dutta, S., Morrison, E.A. and Henzler-Wildman, K.A. 2014. EmrE Dimerization Depends on Membrane Environment. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. **1838**(7), pp.1817–1822.

Efremov, A.K., Yao, M., Sun, Y., Tee, Y.H., Sheetz, M.P., Bershadsky, A.D., Martinac, B. and Yan, J. 2022. Application of piconewton forces to individual filopodia reveals mechanosensory role of L-type Ca2+ channels. *Biomaterials*. **284**, p.121477.

Elazar, A., Weinstein, J.J., Prilusky, J. and Fleishman, S.J. 2016. Interplay Between Hydrophobicity and the Positive-Inside Rule in Determining Membrane-Protein

Topology. *Proceedings of the National Academy of Sciences*. **113**(37), pp.10340–10345.

Evans, E.L., Povstyan, O. V., De Vecchis, D., MacRae, F., Lichtenstein, L., Futers, T.S., Parsonage, G., Humphreys, N.E., Adamson, A., Kalli, A.C., Ludlow, M.J. and Beech, D.J. 2020. RBCs prevent rapid PIEZO1 inactivation and expose slow deactivation as a mechanism of dehydrated hereditary stomatocytosis. *Blood*. **136**(1), pp.140–144.

Fantini, J. and Barrantes, F.J. 2013. How cholesterol interacts with membrane proteins: An exploration of cholesterol-binding sites including CRAC, CARC, and tilted domains. *Frontiers in Physiology*. **4 FEB**.

Ferjani, A., Segami, S., Horiguchi, G., Muto, Y., Maeshima, M. and Tsukaya, H. 2011. Keep an Eye on PP_i: The Vacuolar-Type H^+-Pyrophosphatase Regulates Postgerminative Development in *Arabidopsis*. *The Plant Cell*. **23**(8), pp.2895–2908.

Flegr, J., Prandota, J., Sovičková, M. and Israili, Z.H. 2014. Toxoplasmosis – A Global Threat. Correlation of Latent Toxoplasmosis with Specific Disease Burden in a Set of 88 Countries D. Fernandez-Reyes, ed. *PLoS ONE*. **9**(3), p.e90203.

Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knoll, E.H., Shelley, M., Perry, J.K., Shaw, D.E., Francis, P. and Shenkin, P.S. 2004. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry*. **47**(7), pp.1739–1749.

Galluccio, M., Console, L., Pochini, L., Scalise, M., Giangregorio, N. and Indiveri, C. 2022. Strategies for Successful Over-Expression of Human Membrane Transport Systems Using Bacterial Hosts: Future Perspectives. *International Journal of Molecular Sciences*. **23**(7), p.3823.

García-Nafría, J. and Tate, C.G. 2020. Cryo-Electron Microscopy: Moving Beyond X-Ray Crystal Structures for Drug Receptors and Drug Development. *Annual Review of Pharmacology and Toxicology*. **60**(1), pp.51–71.

Gaxiola, R.A., Palmgren, M.G. and Schumacher, K. 2007. Plant Proton Pumps. *FEBS Letters*. **581**(12), pp.2204–2214.

Ge, J., Li, W., Zhao, Q., Li, N., Chen, M., Zhi, P., Li, R., Gao, N., Xiao, B. and Yang, M. 2015. Architecture of the mammalian mechanosensitive Piezo1 channel. *Nature*. **527**(7576), pp.64–69.

Geng, J., Liu, W., Zhou, H., Zhang, T., Wang, L., Zhang, M., Li, Y., Shen, B., Li, X. and Xiao, B. 2020. A Plug-and-Latch Mechanism for Gating the Mechanosensitive Piezo Channel. *Neuron*. **106**(3), pp.438-451.e6.

Gnanasambandam, R., Bae, C., Gottlieb, P.A. and Sachs, F. 2015. Ionic Selectivity and Permeation Properties of Human PIEZO1 Channels S. Barnes, ed. *PLOS ONE*. **10**(5), p.e0125503.

Goldstein, E.J. 1996. Anaerobic Bacteremia. *Clinical Infectious Diseases*. **23**, pp.S97-101.

Guan, Z., Tian, B., Perfumo, A. and Goldfine, H. 2013. The polar lipids of Clostridium psychrophilum, an anaerobic psychrophile. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*. **1831**(6), pp.1108–1112.

Gubbens, J. and de Kroon, A.I.P.M. 2010. Proteome-Wide Detection of Phospholipid–Protein Interactions in Mitochondria by Photocrosslinking and Click Chemistry. *Molecular BioSystems*. **6**(10), p.1751.

Guo, Y. 2021. Detergent-free systems for structural studies of membrane proteins. *Biochemical Society Transactions*. **49**(3), pp.1316–1374.

Guo, Y.R. and MacKinnon, R. 2017. Structure-based membrane dome mechanism for piezo mechanosensitivity. *eLife*. **6**.

Gupta, K., Donlan, J.A.C., Hopper, J.T.S., Uzdavinys, P., Landreh, M., Struwe, W.B., Drew, D., Baldwin, A.J., Stansfeld, P.J. and Robinson, C. V. 2017. The Role of Interfacial Lipids in Stabilizing Membrane Protein Oligomers. *Nature*. **541**(7637), pp.421–424.

Guvench, O. and MacKerell, A.D. 2008. Comparison of protein force fields for molecular dynamics simulations. *Methods in Molecular Biology*. **443**, pp.63–88.

Habeck, M., Haviv, H., Katz, A., Kapri-Pardes, E., Ayciriex, S., Shevchenko, A., Ogawa, H., Toyoshima, C. and Karlish, S.J.D. 2015. Stimulation, Inhibition, or Stabilization of Na^+,K^+-ATPase Caused by Specific Lipid Interactions at Distinct Sites. *Journal of Biological Chemistry*. **290**(8), pp.4829–4842.

Halgren, T.A., Murphy, R.B., Friesner, R.A., Beard, H.S., Frye, L.L., Pollard, W.T. and Banks, J.L. 2004. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *Journal of Medicinal Chemistry*. **47**(7), pp.1750–1759.

Harborne, S.P.D., Strauss, J., Boakes, J.C., Wright, D.L., Henderson, J.G., Boivineau, J., Jaakola, V.P. and Goldman, A. 2020. IMPROvER: the Integral Membrane Protein

Stability Selector. *Scientific Reports*. **10**(1), pp.1–18.

Hardy, D., Bill, R.M., Jawhari, A. and Rothnie, A.J. 2016. Overcoming bottlenecks in the membrane protein structural biology pipeline. *Biochemical Society Transactions*. **44**(3), pp.838–844.

Haselwandter, C.A. and Mackinnon, R. 2018. Piezo's membrane footprint and its contribution to mechanosensitivity. *eLife*. **7**.

Haviv, H., Cohen, E., Lifshitz, Y., Tal, D.M., Goldshleger, R. and Karlish, S.J.D. 2007. Stabilization of Na^+,K^+-ATPase Purified from *Pichia pastoris* Membranes by Specific Interactions with Lipids. *Biochemistry*. **46**(44), pp.12855–12867.

Haviv, H., Habeck, M., Kanai, R., Toyoshima, C. and Karlish, S.J.D. 2013. Neutral Phospholipids Stimulate Na^+,K^+-ATPase Activity: A Specific Lipid-Protein Interaction. *Journal of Biological Chemistry*. **288**(14), pp.10073–10081.

von Heijne, G. 1992. Membrane Protein Structure Prediction. Hydrophobicity Analysis and the Positive-Inside Rule. *Journal of molecular biology*. **225**(2), pp.487–94.

Heinonen, J.K. 2001. *Biological Role of Inorganic Pyrophosphate* [Online]. Boston, MA: Springer US. [Accessed 24 September 2019]. Available from: http://link.springer.com/10.1007/978-1-4615-1433-6.

Hill, R.Z., Loud, M.C., Dubin, A.E., Peet, B. and Patapoutian, A. 2022. PIEZO1 transduces mechanical itch in mice. *Nature*. **607**(7917), pp.104–110.

Holmes, A.O.M., Goldman, A. and Kalli, A.C. 2022. mPPases create a conserved anionic membrane fingerprint as identified via multi-scale simulations P. M. Kasson, ed. *PLOS Computational Biology*. **18**(10), p.e1010578.

Holmes, A.O.M., Kalli, A.C., Goldman, A. and Goldman, A. 2019. The Function of Membrane Integral Pyrophosphatases From Whole Organism to Single Molecule. *Frontiers in Molecular Biosciences*. **6**(November), pp.1–11.

Holt, J.R., Zeng, W.Z., Evans, E.L., Woo, S.H., Ma, S., Abuwarda, H., Loud, M., Patapoutian, A. and Pathak, M.M. 2021. Spatiotemporal dynamics of piezo1 localization controls keratinocyte migration during wound healing. *eLife*. **10**.

Huang, H., Ge, B., Sun, C., Zhang, S. and Huang, F. 2019. Membrane curvature affects the stability and folding kinetics of bacteriorhodopsin. *Process Biochemistry*. **76**, pp.111–117.

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., De Groot, B.L., Grubmüller, H. and

MacKerell, A.D. 2016. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nature Methods*. **14**(1), pp.71–73.

Huber, R.G., Marzinek, J.K., Holdbrook, D.A. and Bond, P.J. 2017. Multiscale molecular dynamics simulation approaches to the structure and dynamics of viruses. *Progress in Biophysics and Molecular Biology*. **128**, pp.121–132.

Humphrey, W., Dalke, A. and Schulten, K. 1996. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics*. **14**(1), pp.33–38.

Jankovsky, N., Caulier, A., Demagny, J., Guitton, C., Djordjevic, S., Lebon, D., Ouled-Haddou, H., Picard, V. and Garçon, L. 2021. Recent advances in the pathophysiology of PIEZO1-related hereditary xerocytosis. *American Journal of Hematology*. **96**(8), pp.1017–1026.

Jiang, W., Lacroix, J. and Luo, Y.L. 2022. Importance of Molecular Dynamics Equilibrium Protocol on Protein-lipid Interactions near Channel Pore. *bioRxiv*., 2021.11.28.470286.

Jiang, W., Lacroix, J. and Luo, Y.L. 2021. Lipids in the Piezo1 channel pore: gating mechanism, or simulation artifact? *bioRxiv*., 2021.11.28.470286.

Jiang, W., Del Rosario, J.S., Botello-Smith, W., Zhao, S., Lin, Y. chun, Zhang, H., Lacroix, J., Rohacs, T. and Luo, Y.L. 2021. Crowding-induced opening of the mechanosensitive Piezo1 channel in silico. *Communications Biology*. **4**(1).

Jones, D.T. 1999. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology*. **292**(2), pp.195–202.

Jones, G., Willett, P., Glen, R.C., Leach, A.R. and Taylor, R. 1997. Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*. **267**(3), pp.727–748.

De Jong, D.H., Singh, G., Bennett, W.F.D., Arnarez, C., Wassenaar, T.A., Schäfer, L. V., Periole, X., Tieleman, D.P. and Marrink, S.J. 2013. Improved parameters for the martini coarse-grained protein force field. *Journal of Chemical Theory and Computation*. **9**(1), pp.687–697.

Jorgensen, W.L., Maxwell, D.S. and Tirado-Rives, J. 1996. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*. **118**(45), pp.11225–11236.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C.,

Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P. and Hassabis, D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature*. **596**(7873), pp.583–589.

Kajander, T., Kellosalo, J. and Goldman, A. 2013. Inorganic Pyrophosphatases: One Substrate, Three Mechanisms. *FEBS Letters*. **587**(13), pp.1863–1869.

Kalli, A.C. and Reithmeier, R.A.F. 2018. Interaction of the human erythrocyte Band 3 anion exchanger 1 (AE1, SLC4A1) with lipids and glycophorin A: Molecular organization of the Wright (Wr) blood group antigen R. L. Dunbrack, ed. *PLOS Computational Biology*. **14**(7), p.e1006284.

Kalli, A.C., Sansom, M.S.P. and Reithmeier, R.A.F. 2015. Molecular Dynamics Simulations of the Bacterial UraA H^+-Uracil Symporter in Lipid Bilayers Reveal a Closed State and a Selective Interaction with Cardiolipin. *PLoS Computational Biology*. **11**(3), pp.1–27.

Kamajaya, A., Kaiser, J.T., Lee, J., Reid, M. and Rees, D.C. 2014. The structure of a conserved piezo channel domain reveals a topologically distinct β sandwich fold. *Structure*. **22**(10), pp.1520–1527.

Kampjut, D., Steiner, J. and Sazanov, L.A. 2021. Cryo-EM grid optimization for membrane proteins. *iScience*. **24**(3), p.102139.

Kankare, J., Salminen, T., Lahti, R., Cooperman, B.S., Baykov, A.A. and Goldman, A. 1996. Sructure of *Escherichia coli* Inorganic Pyrophosphatase at 2.2 Å Resolution. *Acta Crystallographica Section D Biological Crystallography*. **52**(3), pp.551–563.

Kapsalis, C., Ma, Y., Bode, B.E. and Pliotas, C. 2020. In-Lipid Structure of Pressure-Sensitive Domains Hints Mechanosensitive Channel Functional Diversity. *Biophysical Journal*. **119**(2), pp.448–459.

Kates, M., Syz, J.Y., Gosser, D. and Haines, T.H. 1993. pH-Dissociation Characteristics of Cardiolipin and its 2'-Deoxy Analogue. *Lipids*. **28**(10), pp.877–82.

Kawate, T., Michel, J.C., Birdsong, W.T. and Gouaux, E. 2009. Crystal structure of the ATP-gated P2X4 ion channel in the closed state. *Nature*. **460**(7255), pp.592–598.

Kellenberger, S. and Grutter, T. 2015. Architectural and functional similarities between trimeric ATP-Gated P2X receptors and acid-sensing ion channels. *Journal of Molecular Biology*. **427**(1), pp.54–66.

Kellosalo, J., Kajander, T., Kogan, K., Pokharel, K. and Goldman, A. 2012. The Structure and Catalytic Cycle of a Sodium-Pumping Pyrophosphatase. *Science*. **337**(6093), pp.473–476.

Kempner, E.S. 1993. Damage to Proteins due to the Direct Action of Ionizing Radiation. *Quarterly Reviews of Biophysics*. **26**(1), pp.27–48.

Kermani, A.A. 2021. A guide to membrane protein X-ray crystallography. *FEBS Journal*. **288**(20), pp.5788–5804.

Krebs, M., Beyhl, D., Gorlich, E., Al-Rasheid, K.A.S., Marten, I., Stierhof, Y.-D., Hedrich, R. and Schumacher, K. 2010. Arabidopsis V-ATPase Activity at the Tonoplast is Required for Efficient Nutrient Storage but not for Sodium Accumulation. *Proceedings of the National Academy of Sciences*. **107**(7), pp.3251–3256.

Kriegel, A., Andrés, Z., Medzihradszky, A., Krüger, F., Scholl, S., Delang, S., Patir-Nebioglu, M.G., Gute, G., Yang, H., Murphy, A.S., Peer, W.A., Pfeiffer, A., Krebs, M., Lohmann, J.U. and Schumacher, K. 2015. Job Sharing in the Endomembrane System: Vacuolar Acidification Requires the Combined Activity of V-ATPase and V-PPase. *The Plant Cell*. **27**(12), pp.3383–3396.

Kühlbrandt, W. 2022. Forty years in cryoEM of membrane proteins. *Microscopy*. **71**(S1), pp.I30–I50.

Kwan, T.O.C., Axford, D. and Moraes, I. 2020. Membrane protein crystallography in the era of modern structural biology. *Biochemical Society Transactions*. **48**(6), pp.2505–2524.

Laganowsky, A., Reading, E., Allison, T.M., Ulmschneider, M.B., Degiacomi, M.T., Baldwin, A.J. and Robinson, C. V. 2014. Membrane proteins bind lipids selectively to modulate their structure and function. *Nature*. **510**(7503), pp.172–175.

Landreh, M., Marklund, E.G., Uzdavinys, P., Degiacomi, M.T., Coincon, M., Gault, J., Gupta, K., Liko, I., Benesch, J.L.P., Drew, D. and Robinson, C. V. 2017. Integrating Mass Spectrometry with MD Simulations Reveals the Role of Lipids in Na^+/H^+-Antiporters. *Nature Communications*. **8**, pp.1–9.

Lee, J., Cheng, X., Swails, J.M., Yeom, M.S., Eastman, P.K., Lemkul, J.A., Wei, S., Buckner, J., Jeong, J.C., Qi, Y., Jo, S., Pande, V.S., Case, D.A., Brooks, C.L., MacKerell, A.D., Klauda, J.B. and Im, W. 2016. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation*. **12**(1), pp.405–413.

Lee, S.C., Knowles, T.J., Postis, V.L.G., Jamshad, M., Parslow, R.A., Lin, Y.P., Goldman, A., Sridhar, P., Overduin, M., Muench, S.P. and Dafforn, T.R. 2016. A method for detergent-free isolation of membrane proteins in their local lipid environment. *Nature Protocols*. **11**(7), pp.1149–1162.

Lemercier, G., Dutoya, S., Luo, S., Ruiz, F.A., Rodrigues, C.O., Baltz, T., Docampo, R. and Bakalara, N. 2002. A Vacuolar-Type H^+-Pyrophosphatase Governs Maintenance of Functional Acidocalcisomes and Growth of the Insect and Mammalian Forms of *Trypanosoma brucei*. *Journal of Biological Chemistry*. **277**(40), pp.37369–37376.

Lewis, A.H. and Grandl, J. 2020. Inactivation Kinetics and Mechanical Gating of Piezo1 Ion Channels Depend on Subdomains within the Cap. *Cell Reports*. **30**(3), pp.870-880.e2.

Li, J., Hou, B., Tumova, S., Muraki, K., Bruns, A., Ludlow, M.J., Sedo, A., Hyman, A.J., McKeown, L., Young, R.S., Yuldasheva, N.Y., Majeed, Y., Wilson, L.A., Rode, B., Bailey, M.A., Kim, H.R., Fu, Z., Carter, D.A.L., Bilton, J., Imrie, H., Ajuh, P., Dear, T.N., Cubbon, R.M., Kearney, M.T., Prasad, K.R., Evans, P.C., Ainscough, J.F.X. and Beech, D.J. 2014. Piezo1 integration of vascular architecture with physiological force. *Nature*. **515**(7526), pp.279–282.

Li, J., Yang, H., Peer, W.A., Richter, G., Titapiwantakun, B., Undurraga, S., Murphy, A.S., Gilroy, S. and Gaxiola, R. 2005. Arabidopsis H^+-PPase AVP1 Regulates Auxin-Mediated Organ Development. . **310**(5745), pp.121–125.

Li, K.M., Wilkinson, C., Kellosalo, J., Tsai, J.Y., Kajander, T., Jeuken, L.J.C., Sun, Y.J. and Goldman, A. 2016. Membrane Pyrophosphatases from *Thermotoga maritima* and *Vigna radiata* Suggest a Conserved Coupling Mechanism. *Nature Communications*. **7**(13596), pp.1–11.

Lin, S.M., Tsai, J.Y., Hsiao, C.D., Huang, Y.T., Chiu, C.L., Liu, M.H., Tung, J.Y., Liu, T.H., Pan, R.L. and Sun, Y.J. 2012. Crystal Structure of a Membrane-Embedded H^+-Translocating Pyrophosphatase. *Nature*. **484**(7394), pp.399–403.

Lin, Y.C., Guo, Y.R., Miyagi, A., Levring, J., MacKinnon, R. and Scheuring, S. 2019. Force-induced conformational changes in PIEZO1. *Nature*. **573**(7773), pp.230–234.

Lindahl, E.R. 2008. Molecular dynamics simulations. *Methods in Molecular Biology*. **443**, pp.3–23.

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

Molecular Microbiology. **93**(4), pp.698–712.

Loura, L.M.S., Prieto, M. and Fernandes, F. 2010. The More You See: Spectroscopy in Molecular Biophysics. *European Biophysics Journal*. **39**(4), pp.721–721.

Lukacs, V., Mathur, J., Mao, R., Bayrak-Toydemir, P., Procter, M., Cahalan, S.M., Kim, H.J., Bandell, M., Longo, N., Day, R.W., Stevenson, D.A., Patapoutian, A. and Krock, B.L. 2015. Impaired PIEZO1 function in patients with a novel autosomal recessive congenital lymphatic dysplasia. *Nature Communications*. **6**(1), pp.1–7.

Luoto, H.H., Baykov, A.A., Lahti, R. and Malinen, A.M. 2013. Membrane-Integral Pyrophosphatase Subfamily Capable of Translocating Both Na^+ and H^+. *Proceedings of the National Academy of Sciences*. **110**(4), pp.1255–1260.

Luoto, Heidi H, Belogurov, G.A., Baykov, A.A., Lahti, R. and Malinen, A.M. 2011. Na^+-Translocating Membrane Pyrophosphatases are Widespread in the Microbial World and Evolutionarily Precede H^+-Translocating Pyrophosphatases. *The Journal of biological chemistry*. **286**(24), pp.21633–42.

Luoto, Heidi H., Belogurov, G.A., Baykov, A.A., Lahti, R. and Malinen, A.M. 2011. Na^+-Translocating Membrane Pyrophosphatases are Widespread in the Microbial World and Evolutionarily Precede H^+-Translocating Pyrophosphatases. *Journal of Biological Chemistry*. **286**(24), pp.21633–21642.

Luoto, H.H., Nordbo, E., Baykov, A.A., Lahti, R. and Malinen, A.M. 2013. Membrane Na^+-Pyrophosphatases Can Transport Protons at Low Sodium Concentrations. *Journal of Biological Chemistry*. **288**(49), pp.35126–35137.

Luoto, H.H., Nordbo, E., Malinen, A.M., Baykov, A.A. and Lahti, R. 2015. Evolutionarily Divergent, Na^+-Regulated H^+-Transporting Membrane-Bound Pyrophosphatases. *Biochemical Journal*. **467**(2), pp.281–291.

Lushington, G.H. 2014. Comparative modeling of proteins *In*: *Molecular Modeling of Proteins: Second Edition*., pp.309–330.

Lv, S.L., Lian, L.J., Tao, P.L., Li, Z.X., Zhang, K.W. and Zhang, J.R. 2009. Overexpression of *Thellungiella halophila* H^+-PPase (TsVP) in Cotton Enhances Drought Stress Resistance of Plants. *Planta*. **229**(4), pp.899–910.

Maeshima, M. 2000. Vacuolar H^+ pyrophosphatase. *Biochimica et Biophysica Acta*. **1465**(1), pp.37–51.

Maneshi, M.M., Ziegler, L., Sachs, F., Hua, S.Z. and Gottlieb, P.A. 2018. Enantiomeric Aβ

peptides inhibit the fluid shear stress response of PIEZO1. *Scientific Reports*. **8**(1), p.14267.

Marchesini, N., Luo, S., Rodrigues, C.O., Moreno, S.N. and Docampo, R. 2000. Acidocalcisomes and a Vacuolar H^+-Pyrophosphatase in Malaria Parasites. *The Biochemical journal*. **347**(1), pp.243–53.

Marrink, S.J., Corradi, V., Souza, P.C.T., Ingólfsson, H.I., Tieleman, D.P. and Sansom, M.S.P. 2019. Computational Modeling of Realistic Cell Membranes. *Chemical Reviews*.

Marrink, S.J. and Tieleman, D.P. 2013. Perspective on the Martini Model. *Chemical Society Reviews*. **42**(16), pp.6801–6822.

Marsh, D. 2007. Lateral pressure profile, spontaneous curvature frustration, and the incorporation and conformation of proteins in membranes. *Biophysical Journal*. **93**(11), pp.3884–3899.

Martinac, B., Buechner, M., Delcour, A.H., Adler, J. and Kung, C. 1987. Pressure-sensitive ion channel in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*. **84**(8), pp.2297–2301.

Mitsuda, N., Enami, K., Nakata, M., Takeyasu, K. and Sato, M.H. 2001. Novel Type *Arabidopsis thaliana* H^+-PPase is Localized to the Golgi Apparatus. *FEBS Letters*. **488**(1), pp.29–33.

Monticelli, L., Kandasamy, S., Periole, X., Larson, R., Tieleman, D. and Marrink, S. 2008. {T}he {MARTINI} coarse{-}grained force field{:} {E}xtension to proteins. *J. Chem. Theor. Comp.* **4**, pp.819–834.

Morachevskaya, E., Sudarikova, A. and Negulyaev, Y. 2007. Mechanosensitive channel activity and F-actin organization in cholesterol-depleted human leukaemia cells. *Cell Biology International*. **31**(4 SPEC. ISS.), pp.374–381.

Moreno, S.N.J. and Docampo, R. 2009. The Role of Acidocalcisomes in Parasitic Protists. *Journal of Eukaryotic Microbiology*. **56**(3), pp.208–213.

Murphy, E.J., Joseph, L., Stephens, R. and Horrocks, L.A. 1992. Phospholipid composition of cultured human endothelial cells. *Lipids*. **27**(2), pp.150–153.

Nonomura, K., Lukacs, V., Sweet, D.T., Goddard, L.M., Kanie, A., Whitwam, T., Ranade, S.S., Fujimori, T., Kahn, M.L. and Patapoutian, A. 2018. Mechanically activated ion channel PIEZO1 is required for lymphatic valve formation. *Proceedings of the National Academy of Sciences of the United States of America*. **115**(50), pp.12817–12822.

Nordbo, E., Luoto, H.H., Baykov, A.A., Lahti, R. and Malinen, A.M. 2016. Two Independent Evolutionary Routes to Na^+/H^+ Cotransport Function in Membrane Pyrophosphatases. *Biochemical Journal*. **473**(19), pp.3099–3111.

Norimatsu, Y., Hasegawa, K., Shimizu, N. and Toyoshima, C. 2017. Protein-Phospholipid Interplay Revealed with Crystals of a Calcium Pump. *Nature*. **545**(7653), pp.193–198.

Nugent, T. and Jones, D.T. 2012. Detecting pore-lining regions in transmembrane protein sequences. *BMC Bioinformatics*. **13**(1), p.169.

Olofsson, G. and Sparr, E. 2013. Ionization Constants pK_a of Cardiolipin M. Gasset, ed. *PLoS ONE*. **8**(9), p.e73040.

Overington, J.P., Al-Lazikani, B. and Hopkins, A.L. 2006. How many drug targets are there? *Nature Reviews Drug Discovery*. **5**(12), pp.993–996.

Park, S., Li, J., Pittman, J.K., Berkowitz, G.A., Yang, H., Undurraga, S., Morris, J., Hirschi, K.D. and Gaxiola, R.A. 2005. Up-Regulation of a H^+-Pyrophosphatase (H^+-PPase) as a Strategy to Engineer Drought-Resistant Crop Plants. *Proceedings of the National Academy of Sciences*. **102**(52), pp.18830–18835.

Pasapula, V., Shen, G., Kuppu, S., Paez-Valencia, J., Mendoza, M., Hou, P., Chen, J., Qiu, X., Zhu, L., Zhang, X., Auld, D., Blumwald, E., Zhang, H., Gaxiola, R. and Payton, P. 2011. Expression of an Arabidopsis Vacuolar H^+-Pyrophosphatase Gene (AVP1) in Cotton Improves Drought- and Salt Tolerance and Increases Fibre Yield in the Field Conditions. *Plant Biotechnology Journal*. **9**(1), pp.88–99.

Pavić, A., Holmes, A.O.M., Postis, V.L.G. and Goldman, A. 2019. Glutamate transporters: a broad review of the most recent archaeal and human structures. *Biochemical Society transactions*., p.BST20190316.

Peng, T., Yuan, X. and Hang, H.C. 2014. Turning the Spotlight on Protein–Lipid Interactions in Cells. *Current Opinion in Chemical Biology*. **21**, pp.144–153.

Peyronnet, R., Martins, J.R., Duprat, F., Demolombe, S., Arhatte, M., Jodar, M., Tauc, M., Duranton, C., Paulais, M., Teulon, J., Honoré, E. and Patel, A. 2013. Piezo1-dependent stretch-activated channels are inhibited by Polycystin-2 in renal tubular epithelial cells. *EMBO Reports*. **14**(12), pp.1143–1148.

Phillips, R., Ursell, T., Wiggins, P. and Sens, P. 2009. Emerging roles for lipids in shaping membrane-protein function. *Nature*. **459**(7245), pp.379–385.

Pliotas, C., Dahl, A.C.E., Rasmussen, T., Mahendran, K.R., Smith, T.K., Marius, P., Gault,

J., Banda, T., Rasmussen, A., Miller, S., Robinson, C. V., Bayley, H., Sansom, M.S.P., Booth, I.R. and Naismith, J.H. 2015. The role of lipids in mechanosensation. *Nature Structural and Molecular Biology*. **22**(12), pp.991–998.

Pomorski, T.G., Nylander, T. and Cárdenas, M. 2014. Model cell membranes: Discerning lipid and protein contributions in shaping the cell. *Advances in Colloid and Interface Science*. **205**, pp.207–220.

Protein DataBank 2022. PDB Statistics: Number of Released PDB Structures per Year. [Accessed 25 October 2022]. Available from: https://www.rcsb.org/stats/all-released-structures.

Pyle, E., Kalli, A.C., Amillis, S., Hall, Z., Lau, A.M., Hanyaloglu, A.C., Diallinas, G., Byrne, B. and Politis, A. 2018. Structural Lipids Enable the Formation of Functional Oligomers of the Eukaryotic Purine Symporter UapA. *Cell Chemical Biology*. **25**(7), pp.840-848.e4.

Qi, Y., Andolfi, L., Frattini, F., Mayer, F., Lazzarino, M. and Hu, J. 2015. Membrane stiffening by STOML3 facilitates mechanosensation in sensory neurons. *Nature Communications*. **6**(1), pp.1–13.

Quistgaard, E.M., Löw, C., Guettou, F. and Nordlund, P. 2016. Understanding transport by the major facilitator superfamily (MFS): Structures pave the way. *Nature Reviews Molecular Cell Biology*. **17**(2), pp.123–132.

Renaud, J.P., Chari, A., Ciferri, C., Liu, W.T., Rémigy, H.W., Stark, H. and Wiesmann, C. 2018. Cryo-EM in drug discovery: Achievements, limitations and prospects. *Nature Reviews Drug Discovery*. **17**(7), pp.471–492.

Ridone, P., Pandzic, E., Vassalli, M., Cox, C.D., Macmillan, A., Gottlieb, P.A. and Martinac, B. 2020. Disruption of membrane cholesterol organization impairs the activity of PIEZO1 channel clusters. *Journal of General Physiology*. **152**(8).

Rodrigues, C.O., Scott, D.A., Bailey, B.N., De Souza, W., Benchimol, M., Moreno, B., Urbina, J.A., Oldfield, E. and Moreno, S.N. 2000. Vacuolar Proton Pyrophosphatase Activity and Pyrophosphate (PP_i) in *Toxoplasma gondii* as Possible Chemotherapeutic Targets. *The Biochemical journal*. **349**(2000), pp.737–45.

Romero, L.O., Massey, A.E., Mata-Daboin, A.D., Sierra-Valdez, F.J., Chauhan, S.C., Cordero-Morales, J.F. and Vásquez, V. 2019. Dietary fatty acids fine-tune Piezo1 mechanical response. *Nature Communications*. **10**(1), pp.1–14.

Rosenbaum, D.M., Cherezov, V., Hanson, M.A., Rasmussen, S.G.F., Foon, S.T., Kobilka,

T.S., Choi, H.J., Yao, X.J., Weis, W.I., Stevens, R.C. and Kobilka, B.K. 2007. GPCR engineering yields high-resolution structural insights into β2-adrenergic receptor function. *Science*. **318**(5854), pp.1266–1273.

Ruiz, F.A., Rodrigues, C.O. and Docampo, R. 2001. Rapid Changes in Polyphosphate Content within Acidocalcisomes in Response to Cell Growth, Differentiation, and Environmental Stress in *Trypanosoma cruzi*. *Journal of Biological Chemistry*. **276**(28), pp.26114–26121.

Saotome, K., Murthy, S.E., Kefauver, J.M., Whitwam, T., Patapoutian, A. and Ward, A.B. 2018. Structure of the mechanically activated ion channel Piezo1. *Nature*. **554**(7693), pp.481–486.

Sarafian, V., Potier, M. and Poole, R.J. 1992. 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**, pp.493–497.

Sathappa, M. and Alder, N.N. 2016. The Ionization Properties of Cardiolipin and its Variants in Model Bilayers. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. **1858**(6), pp.1362–1372.

Scott, D.A., de Souza, W., Benchimol, M., Zhong, L., Lu, H.G., Moreno, S.N. and Docampo, R. 1998. Presence of a Plant-Like Proton-Pumping Pyrophosphatase in Acidocalcisomes of *Trypanosoma cruzi*. *The Journal of biological chemistry*. **273**(34), pp.22151–8.

Segami, S., Asaoka, M., Kinoshita, S., Fukuda, M., Nakanishi, Y. and Maeshima, M. 2018. Biochemical, structural and physiological characteristics of vacuolar h+-pyrophosphatase. *Plant and Cell Physiology*. **59**(7), pp.1300–1308.

Segami, S., Makino, S., Miyake, A., Asaoka, M. and Maeshima, M. 2014. Dynamics of Vacuoles and H^+-Pyrophosphatase Visualized by Monomeric Green Fluorescent Protein in Arabidopsis: Artifactual Bulbs and Native Intravacuolar Spherical Structures. *The Plant Cell*. **26**(8), pp.3416–3434.

Shah, N., Vidilaseris, K., Xhaard, H. and Goldman, A. 2016. Integral Membrane Pyrophosphatases: A Novel Drug Target for Human Pathogens? *AIMS Biophysics*. **3**(1), pp.171–194.

Shah, N.R., Wilkinson, C., Harborne, S.P.D., Turku, A., Li, K.M., Sun, Y.J., Harris, S. and Goldman, A. 2017. Insights into the Mechanism of Membrane Pyrophosphatases by Combining Experiment and Computer Simulation. *Structural Dynamics*. **4**(3), pp.1–12.

Skach, W.R. 2009. Cellular mechanisms of membrane protein folding. *Nature Structural and Molecular Biology*. **16**(6), pp.606–612.

Sohlenkamp, C. and Geiger, O. 2015. Bacterial membrane lipids: Diversity in structures and pathways. *FEMS Microbiology Reviews*. **40**(1), pp.133–159.

Solis, A.G., Bielecki, P., Steach, H.R., Sharma, L., Harman, C.C.D., Yun, S., de Zoete, M.R., Warnock, J.N., To, S.D.F., York, A.G., Mack, M., Schwartz, M.A., Dela Cruz, C.S., Palm, N.W., Jackson, R. and Flavell, R.A. 2019. Mechanosensation of cyclical force by PIEZO1 is essential for innate immunity. *Nature*. **573**(7772), pp.69–74.

Song, S., Zhang, H., Wang, X., Chen, W., Cao, W., Zhang, Z. and Shi, C. 2022. The role of mechanosensitive Piezo1 channel in diseases. *Progress in Biophysics and Molecular Biology*. **172**, pp.39–49.

Song, W., Corey, R.A., Ansell, T.B., Cassidy, C.K., Horrell, M.R., Duncan, A.L., Stansfeld, P.J. and Sansom, M.S.P. 2022. PyLipID: A Python Package for Analysis of Protein-Lipid Interactions from Molecular Dynamics Simulations. *Journal of Chemical Theory and Computation*. **18**(2), pp.1188–1201.

Song, Y., Dimaio, F., Wang, R.Y.R., Kim, D., Miles, C., Brunette, T., Thompson, J. and Baker, D. 2013. High-resolution comparative modeling with RosettaCM. *Structure*. **21**(10), pp.1735–1742.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E. and Berendsen, H.J.C. 2005. GROMACS: Fast, Flexible, and Free. *Journal of Computational Chemistry*. **26**(16), pp.1701–1718.

Stansfeld, P.J. and Sansom, M.S.P. 2011. Molecular Simulation Approaches to Membrane Proteins. *Structure*. **19**(11), pp.1562–1572.

Strauss, J. 2021. *Towards more rational approaches of membrane protein stabilisation and novel structures of membrane-bound pyrophosphatase*.

Studer, G., Biasini, M. and Schwede, T. 2014. Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane). *Bioinformatics*. **30**(17), pp.i505–i511.

Sukharev, S.I., Martinac, B., Arshavsky, V.Y. and Kung, C. 1993. Two types of mechanosensitive channels in the Escherichia coli cell envelope: solubilization and functional reconstitution. *Biophysical Journal*. **65**(1), pp.177–183.

Sweadner, K.J. 2017. An Ion-Transport Enzyme That Rocks. *Nature*. **545**(7653), pp.162–

164.

Tang, H., Zeng, R., He, E., Zhang, I., Ding, C. and Zhang, A. 2022. Piezo-Type Mechanosensitive Ion Channel Component 1 (Piezo1): A Promising Therapeutic Target and Its Modulators. *Journal of Medicinal Chemistry*. **65**(9), pp.6441–6453.

Tipper, D.J. and Strominger, J.L. 1965. Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-D-alanine. *Proceedings of the National Academy of Sciences of the United States of America*. **54**(4), pp.1133–1141.

Torgerson, P.R. and Mastroiacovo, P. 2013. The Global Burden of Congenital Toxoplasmosis: A Systematic Review. *Bulletin of the World Health Organization*. **91**(7), pp.501–508.

Tsai, J.-Y., Tang, K.-Z., Li, K.-M., Hsu, B.-L., Chiang, Y.-W., Goldman, A. and Sun, Y.-J. 2019. Roles of the Hydrophobic Gate and Exit Channel in *Vigna radiata* Pyrophosphatase Ion Translocation. *Journal of Molecular Biology*. **431**(8), pp.1619–1632.

Tsai, J.Y., Kellosalo, J., Sun, Y.J. and Goldman, A. 2014. Proton/Sodium Pumping Pyrophosphatases: The Last of the Primary Ion Pumps. *Current Opinion in Structural Biology*. **27**(1), pp.38–47.

Tzeng, C.M., Yang, C.Y., Yang, S.J., Jiang, S.S., Kuo, S.Y., Hung, S.-H., Ma, J.T. and Pan, R.L. 1996. Subunit Structure of Vacuolar Proton-Pyrophosphatase as Determined by Radiation Inactivation. *Biochemical Journal*. **316**(1), pp.143–147.

Ursell, T., Kondev, J., Reeves, D., Wiggins, P.A. and RobPhillips, R. 2008. Role of Lipid Bilayer Mechanics in Mechanosensation *In*: *Mechanosensitive Ion Channels* [Online]. Springer Netherlands, pp.37–70. [Accessed 17 October 2022]. Available from: https://link.springer.com/chapter/10.1007/978-1-4020-6426-5_2.

Vanegas, J.M. and Arroyo, M. 2014. Force Transduction and Lipid Binding in MscL: A Continuum-Molecular Approach D. Fotiadis, ed. *PLoS ONE*. **9**(12), p.e113947.

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Zídek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., Cowie, A., Hobbs, N., Kohli, P., Kleywegt, G., Birney, E., Hassabis, D. and Velankar, S. 2022. AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*. **50**(D1), pp.D439–D444.

De Vecchis, D., Beech, D.J. and Kalli, A.C. 2021. Molecular dynamics simulations of Piezo1 channel opening by increases in membrane tension. *Biophysical Journal*. **120**(8), pp.1510–1521.

Vidilaseris, K., Kellosalo, J. and Goldman, A. 2018. A High-Throughput Method for Orthophosphate Determination of Thermostable Membrane-Bound Pyrophosphatase Activity. *Analytical Methods*. **10**(6), pp.646–651.

Vidilaseris, K., Kiriazis, A., Turku, A., Khattab, A., Johansson, N.G., Leino, T.O., Kiuru, P.S., Boije af Gennäs, G., Meri, S., Yli-Kauhaluoma, J., Xhaard, H. and Goldman, A. 2019. Asymmetry in Catalysis by *Thermotoga maritima* Membrane-Bound Pyrophosphatase Demonstrated by a Nonphosphorus Allosteric Inhibitor. *Science Advances*. **5**(5), p.eaav7574.

Volkers, L., Mechioukhi, Y. and Coste, B. 2015. Piezo channels: from structure to function. *Pflugers Archiv European Journal of Physiology*. **467**(1), pp.95–99.

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

Wang, L., Zhou, H., Zhang, M., Liu, W., Deng, T., Zhao, Q., Li, Y., Lei, J., Li, X. and Xiao, B. 2019. Structure and mechanogating of the mammalian tactile channel PIEZO2. *Nature*. **573**(7773), pp.225–229.

Wang, Y., Chi, S., Guo, H., Li, G., Wang, L., Zhao, Q., Rao, Y., Zu, L., He, W. and Xiao, B. 2018. A lever-like transduction pathway for long-distance chemical- and mechano-gating of the mechanosensitive Piezo1 channel. *Nature Communications*. **9**(1), pp.1–12.

Wassenaar, T.A., Ingólfsson, H.I., Böckmann, R.A., Tieleman, D.P. and Marrink, S.J. 2015. Computational lipidomics with insane: A versatile tool for generating custom membranes for molecular simulations. *Journal of Chemical Theory and Computation*. **11**(5), pp.2144–2155.

Wassenaar, T.A., Pluhackova, K., Böckmann, R.A., Marrink, S.J. and Tieleman, D.P. 2014. Going backward: A flexible geometric approach to reverse transformation from coarse grained to atomistic models. *Journal of Chemical Theory and Computation*. **10**(2), pp.676–690.

Webb, B. and Sali, A. 2016. Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics*. **54**(1), pp.1–37.

Wexler, H.M. 2007. Bacteroides: The Good, the Bad, and the Nitty-Gritty. *Clinical Microbiology Reviews*. **20**(4), pp.593–621.

Wilkinson, A.R. 2019. Design, Synthesis and Biological Evaluation of Novel Membrane-Integral Pyrophosphatase Inhibitors.

Woodward, X., Stimpson, E.E. and Kelly, C. V. 2018. Single-lipid tracking on nanoscale membrane buds: The effects of curvature on lipid diffusion and sorting. *Biochimica et Biophysica Acta - Biomembranes*. **1860**(10), pp.2064–2075.

Woolf, T.B. and Roux, B. 1994. Molecular dynamics simulation of the gramicidin channel in a phospholipid bilayer. *Proceedings of the National Academy of Sciences of the United States of America*. **91**(24), pp.11631–11635.

World Health Organization. Global Malaria Programme. 2018. *World Malaria Report 2017.*

Wu, J., Goyal, R. and Grandl, J. 2016. Localized force application reveals mechanically sensitive domains of Piezo1. *Nature Communications*. **7**(1), pp.1–10.

Wu, J.J., Ma, J.T. and Pan, R.L. 1991. Functional Size Analysis of Pyrophosphatase from *Rhodospirillum rubrum* Determined by Radiation Inactivation. *FEBS Letters*. **283**(1), pp.57–60.

Xiao, B. 2020. Levering mechanically activated piezo channels for potential pharmacological intervention. *Annual Review of Pharmacology and Toxicology*. **60**, pp.195–218.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. and Zhang, Y. 2014. The I-TASSER suite: Protein structure and function prediction. *Nature Methods*. **12**(1), pp.7–8.

Yang, S.J., Jiang, S.S., Van, R.C., Hsiao, Y.Y. and Pan, R.L. 2000. A Lysine Residue Involved in the Inhibition of Vacuolar H^+-Pyrophosphatase by Fluorescein 5'-Isothiocyanate. *Biochimica et Biophysica Acta - Bioenergetics*. **1460**(2), pp.375–383.

Yang, X., Lin, C., Chen, X., Li, S., Li, X. and Xiao, B. 2022. Structure deformation and curvature sensing of PIEZO1 in lipid membranes. *Nature*. **604**(7905), pp.377–383.

Yeagle, P.L. 2014. Non-Covalent Binding of Membrane Lipids to Membrane Proteins. *Biochimica et Biophysica Acta - Biomembranes*. **1838**(6), pp.1548–1559.

Yoon, H.S., Kim, S.Y. and Kim, I.S. 2013. Stress Response of Plant H^+-PPase-Expressing Transgenic *Escherichia coli* and *Saccharomyces cerevisiae*: A Potentially Useful Mechanism for the Development of Stress-Tolerant Organisms. *Journal of Applied Genetics*. **54**(1), pp.129–133.

Yoshida, S. and Uemura, M. 1986. Lipid Composition of Plasma Membranes and Tonoplasts Isolated from Etiolated Seedlings of Mung Bean (*Vigna radiata* L.). *Plant Physiology*. **82**(3), pp.807–812.

Zhang, H., Shen, G., Kuppu, S., Gaxiola, R. and Payton, P. 2011. Creating Drought- and Salt-Tolerant Cotton by Overexpressing a Vacuolar Pyrophosphatase Gene. *Plant Signaling and Behavior*. **6**(6), pp.861–863.

Zhang, M., Wang, C., Otto, T.D., Oberstaller, J., Liao, X., Adapa, S.R., Udenze, K., Bronner, I.F., Casandra, D., Mayho, M., Brown, J., Li, S., Swanson, J., Rayner, J.C., Jiang, R.H.Y. and Adams, J.H. 2018. Uncovering the Essential Genes of the Human Malaria Parasite *Plasmodium falciparum* by Saturation Mutagenesis. *Science (New York, N.Y.)*. **360**(6388), p.eaap7847.

Zhao, F.-Y., Zhang, X.-J., Li, P.-H., Zhao, Y.-X. and Zhang, H. 2006. Co-Expression of the *Suaeda salsa* SsNHX1 and Arabidopsis AVP1 Confer Greater Salt Tolerance to Transgenic Rice than the Single SsNHX1. *Molecular Breeding*. **17**(4), pp.341–353.

Zhao, Q., Wu, K., Geng, J., Chi, S., Wang, Y., Zhi, P., Zhang, M. and Xiao, B. 2016. Ion Permeation and Mechanotransduction Mechanisms of Mechanosensitive Piezo Channels. *Neuron*. **89**(6), pp.1248–1263.

Zhao, Q., Zhou, H., Chi, S., Wang, Y., Wang, Jianhua, Geng, J., Wu, K., Liu, W., Zhang, T., Dong, M.Q., Wang, Jiawei, Li, X. and Xiao, B. 2018. Structure and mechanogating mechanism of the Piezo1 channel. *Nature*. **554**(7693), pp.487–492.

Zhao, Q., Zhou, H., Li, X. and Xiao, B. 2019. The mechanosensitive Piezo1 channel: a three-bladed propeller-like structure and a lever-like mechanogating mechanism. *The FEBS Journal*. **286**(13), pp.2461–2470.