Getting into the groove! Investigating how lipids, ions and water interact with dynamic protein cavities using physics-based simulations

by
Christina Stephens

THESIS
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biophysics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Michael Grabe
5B8B990D587B414...

Michael Grabe
Chair

DocuSigned by:
Matthew Jacobson

Matthew Jacobson

DocuSigned by:
Lily Jan
394456D87457494...

Lily Jan

Committee Members

DEDICATION AND ACKNOWLEDGEMENTS

This work is dedicated to my parents, my sisters, and my husband. None of this would have been possible without their love and unwavering support.

I wish to dedicate the last chapter of this thesis specifically to Dr. Marcel Bruchez who was a brilliant and devoted scientist. It was a pleasure to work with him and a privilege to learn from him.

I also want to acknowledge here the community at UCSF that has supported me intellectually throughout this journey. I'm grateful for the countless insightful conversations with my advisor, Dr. Michael Grabe, who has always managed to make our science fun and exciting. I also thank the rest of my committee, Dr. Lily Jan and Dr. Matthew Jacobson, for enriching my knowledge and keeping me on track. I've been fortunate to have learned so much from other members of the Grabe lab as well, particularly Andrew Natale, Yessica Gomez, and Paola Bisignano. It has also been so enjoyable working with Niek van Hilten and Lisa Zheng on our recent projects. I want to thank my friends who have supported me in and out of the lab: Nadia, Matt, Calla, Wren, Maru, Pooja, Justin, Jinhua, Matvei, Yessica (again), Andrew (again), Wade, and Sydney. Finally, I'd like to thank my previous advisors Dr. Lucy Forrest, Dr. Shantá Hinton and Dr. Randolph Coleman for teaching what a PhD means but not telling me how hard it would be.

Research Acknowledgments: Andrew Natale (density & membrane deformation code), Yessica Gomez (plotting membrane deformations), Paola Bisignano (loop modelling 16F), George Khelashvili (sharing simulated structure open TMEM16F). Preliminary computational research for work presented in Chapter 4 performed by Chiara Blake.

CONTRIBUTIONS

Chapter 2 has been adapted from a manuscript in preparation for submission. Co-authors on this work include Christian Peters and Michael Grabe.

Chapter 3 has been adapted from a manuscript in preparation for submission. Co-authors on this work include Niek van Hilten, Lisa Zheng and Michael Grabe. Niek van Hilten and Lisa Zheng made writing and research contributions to parts of the results.

Chapter 4 has been adapted from a manuscript in preparation for submission. Co-authors on this work include Huaxia Shi, Peng Zhang, Marcel Bruchez, David Beratan and Michael Grabe.

GETTING INTO THE GROOVE! INVESTIGATING HOW LIPIDS, IONS AND WATER INTERACT WITH DYNAMIC PROTEIN CAVITIES USING PHYSICS-BASED SIMULATIONS

Christina Alexandra Stephens

ABSTRACT

Proteins, both membrane dwelling and soluble, have the remarkable ability to alter their shapes when they interact with other molecules. My work has focused on the ability of one family of membrane proteins and a completely unrelated soluble protein to open and close entrances to cavities and grooves in ways that are fundamental to how they perform their functions. In the first instance, members of the TMEM16 family of Ca^{2+}-activated ion channels and lipid scramblases open a hydrophilic channel or "groove" by which solvent molecules move passively down their electrochemical gradients. However, there is currently ongoing debate over their precise transport mechanisms including whether an open groove is necessary for lipid scrambling, the role of membrane deformation and finally how voltage and lipids modulate TMEM16 activity. I will first describe how I have used atomistic molecular dynamics (MD) simulations to predict new ion-conductive states of the TMEM16A anion channel and show how volage-dependent rotamers of basic residues give rise to current rectification observed in experiments. Then I will describe how I have used an enhanced sampling MD technique to survey 27 TMEM16 scramblase and ion channel structures and determined that these family members do require both an open groove and significant membrane deformation to scramble lipids. My analysis of TMEM16A from both sets of simulations also reveals that lipids partially form the ion conduction pathway without being transported themselves.

The latter instance concerns a symmetric light chain antibody dimer that binds heavy atom substituted fluorescent dyes and produces singlet oxygen upon light excitation only when the dye is bound to the

protein complex dimer interface. Analogs of the dye molecule give rise to a wide range in the singlet oxygen quantum yields, but it is not understood what mechanism underlies this phenomenon. In the final part of this thesis, I will describe how I have used atomistic MD simulations to discuss how dynamics of the varying part of the dye analogs may be coupled to a conformation change of the protein that widens the dimer cleft, alters solvent exposure to the dye itself and thereby allows excited dye molecules to transfer their energy to molecular oxygen.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

$^1O_2$ QY – singlet oxygen quantum yield

1PBC - hydroxy-3-(trifluoromethyl)pyrido[1,2-a]benzimidazole-4-carbonitrile

1PBC(-1) – negatively charged 1PBC

1PBC(0) – neutrally charged (protonated) 1PBC

Å – Angstrom

AAMD – all-atom molecular dynamics

Ano – Anoctamin

atm – atmosphere

Ca^{2+} - calcium ion

CaCC – Calcium-activated chloride channel

CGMD – coarse-grained molecular dynamics

Cl^- - chloride ion

cryo-EM - cryogenic electron microscopy

Cα – alpha carbon

e – electron charge

ER – Endoplasmic Reticulum

fs - femtosecond

G – conductance

k – Boltzmann constant

K – kelvin

K^+ - potassium ion

kcal – kilocalorie

kJ – kilojoule

LINCS – LINear Constraint Solver

MG – malachite green

mM – millimolar

MO – molecular oxygen

MSM – Markov state model

mTMEM16A – mouse TMEM16

mV - millivolt

Na^+ - sodium

nm - nanometer

ns – nanosecond

pA – picoampere

PIP_2 – phosphatidylinositol-4,5-bisphosphate

POPC - 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine

ps - picosecond

pS – picosiemen

QM – quantum mechanics

RMSD – root-mean-squared deviation

S0 – ground electronic singlet state

S1 – first electronic excited singlet state

T – temperature (in kelvin)

T1 – firs electronic excited triplet state

tICA – time-independent component analysis

TM – transmembrane (helix)

TMEM16 – TransMEMbrane protein 16

VdW - Van der Waals

x-Ray – x-ray crystallography

μM – micromolar

μs – microsecond

1. BACKGROUND

Cells, the fundamental building blocks of life, require a specific internal balance of ions and asymmetric distribution of lipids in their plasma membranes to regulate their pH, pressure, membrane shape, and normal day-to-day molecular signaling. They do this by expending energy, typically by breaking bonds in ATP molecules, to pump ions and lipids against their electrochemical gradients. In turn cells can also harvest these gradients to do work, for example export sugar out the cell by coupling its transport to the favorable movement of Na^+ down its gradient [1] or induce largescale membrane rearrangements by rapidly collapsing leaflet asymmetries [2]. A wide variety of proteins that sit in cell membranes, both internally such as in the ER membrane and at the plasma membrane, are responsible moving substrates in this passive way. The first two chapters of this thesis will focus of a family of membrane proteins that provide a conduit for lipids and ions to move passively down their gradients in response to specific electrical and chemical stimuli. In the field we refer to membrane proteins that allow the bidirectionally flow of lipids to opposing leaflets as scramblases.

The TMEM16 Family of Lipid Scramblases and Ion Channels

The TMEM16 family of eukaryotic membrane proteins, also known as anoctamins, Ano, is comprised of lipid scramblases [3][4][5], ion channels [6][7][8][9], and members that can facilitate both lipid and ion permeation [10][11][12][13][14][15][16][17]. This functional divergence, despite their high sequence conservation, is a unique feature among the 10 vertebrate paralogues [18]. So far, all characterized TMEM16s require Ca^{2+} to achieve their maximum transport activity, that is enable the passive movement of ions or ions and lipids down their electrochemical gradients [19], [20][21][22][23]. TMEM16s play critical roles in a variety of physiological processes including blood coagulation [10][24][25][26], bone mineralization [27], mucus secretion [28], smooth muscle contraction [29] and membrane fusion [30]. Mutations to TMEM16 have also been implicated in several cancers [31][32][33],

neuronal disorder SCAR10 [34][35], and SARS-CoV2 infection [36]. Despite their significant roles in human physiology, the functional properties of most vertebrate TMEM16 paralogues have remained unclear. Moreover, even though we have significant functional and structural insight into the mechanisms of a handful of members [13][15][17][37][38][39][40][41][42][43][44][45] [46], [47][48][49], [50][51][52][53], [54][55][56][57][58][59][60] it is still an open question whether all TMEM16s work in the same way.

Over the past 10 years 63 experimental structures of TMEM16s have been determined exposing a remarkable structural similarity between both mammalian and fungal members despite the diversity in their functions. All structures, except for one of fungal *Aspergillus fumigatus* TMEM16 (afTMEM16) [61], are of homodimers with a butterfly-like fold [14][17], [21][23][62][63][64][65][66][67][68], [69], [70][71][72] and each subunit is comprised of 10 transmembrane (TM) helices with the final helix (TM10) forming most of the dimer interface. Residues on TM6 form half of a highly conserved Ca^{2+}-binding site that accommodates up to 2 ions. TM6 along with 3, 4, and 5 also form a membrane spanning groove that contains hydrophilic residues that are shielded from the hydrophobic core of the bilayer in Ca^{2+}-free states. When Ca^{2+} is bound, TM6 takes on a variety of conformation and secondary structure changes across the family (**Figure B.1**) which can have profound effects on the shape of the membrane as seen in cryo-EM nanodiscs with TMEM16F [71]. Ca^{2+}-binding is also associated with movement of upper TM4 away from TM6 which effectively exposes (opens) the hydrophilic groove, but this is not observed for all TMEM16 members [65] [23], [71], [72] [67][48][68], [69], [70]. For example, no structures of TMEM16A, which is an ion channel and does not scramble lipids, has this movement of TM4 away from TM6, at least to the same extent as say nhTMEM16. However, TMEM16A does take on a 45-60° rotation of its TM3, another helix lining putative ion pore, in some Ca^{2+}-bound structures, when either bound to an inhibitor (**Figure 1.1**) or in the WT C-terminal-truncated highly voltage sensitive isoform. Importantly, this rotation brings a basic residue, known to have a functional role in ion selectivity [46], into the putative pore pathway but so far there has not been a thorough investigation into the electrophysiological

properties of this TM3-rotated state at the atomistic level. The second chapter of this thesis focuses on my efforts to characterize the electronic properties of this TMEM16A conformation while chapter 3 will take a broader look at which TMEM16 structural features are critical for their ability to scramble.

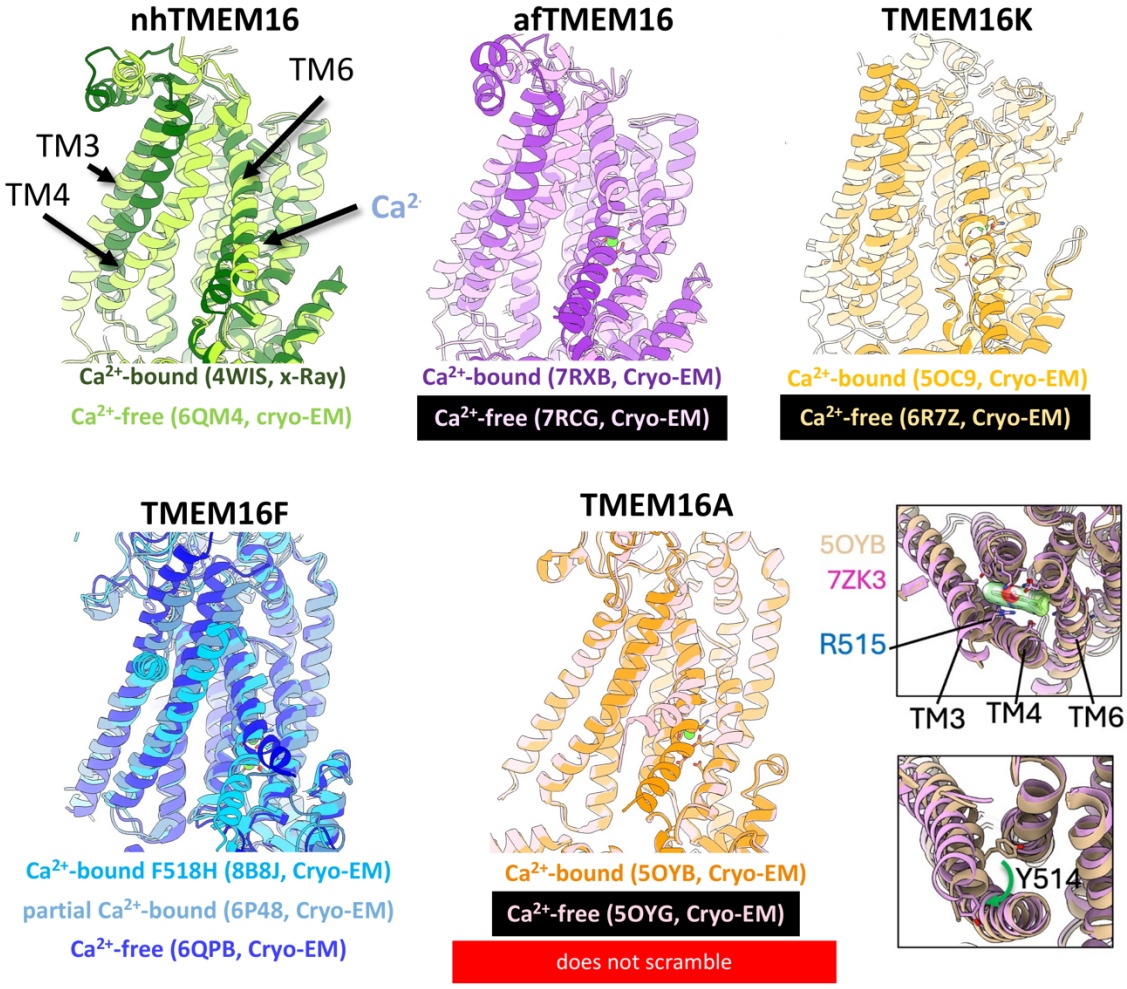

Figure 1.1 Structural comparison of experimentally solved TMEM16 structures.
Superposition of single subunits from dimer TMEM16 proteins: PDB IDs: 4WIS [21], 6QM4 [65], 7RXB [62], 7RXG [62], 5OC9 [14], 6R7Z [14], 8B8J [64], 6P48 [71], 6QPB [23], 5OYB [63], 5OYG [63], 7ZK3 [51].

Molecular Dynamics Simulations

One way to probe the atomistic mechanisms of molecules such as these membrane proteins is to simulate their movement over a short (ns-μs) timescale. Atomistic molecular dynamics (MD) simulation is a powerful computational method that uses classical physics to represent the potentials and force felt by

every atom in an enclosed system and predict their motion over time. In this type of simulation, the atoms are represented as hard spheres and all bonds between atoms are fixed. The atomic potentials include contributions from bonded (e.g. characteristic bond lengths represented by spring constants) and nonbonded interactions (i.e. coulombic, VdW and electronic) which are pulled from a set of parameters for every type of atom in a library referred to as a forcefield. These parameters have been optimized to either match the energic landscape of *ab initio* quantum mechanical calculations and experimental observations such a lipid diffusion rates. For the purposes of studying the dynamics of proteins the simulation systems are set up to mimic their natural biological environment. In the case of membrane proteins this means embedding the protein in a lipid bilayer (although this is often simplified to one or several lipid species) and packing the area below and above the bilayer with solvent, typically water and a physiological concentration of ions (**Figure 1.2**). Representing the system at the atomistic level can provided deep insight into how the protein interacts with other molecules such as lipids, drugs and even other proteins and how they alter their shapes in response to different stimuli. I have used all-atom MD (AAMD) simulations to successfully sample major conformational changes of TMEM16A and a soluble protein-ligand complex which are detailed in Chapters 2 and 4 of this thesis. But such a detailed representation of the system can be computationally costly; for example, it can take several months to collect a few μs of simulation time for a system of around 300,000 atoms. This method becomes intractable when one's goal is to observe multiple events that either happen over longer time scales or occur rarely within that time scale. Lipid and ion transport by TMEM16s is a great example of the latter.

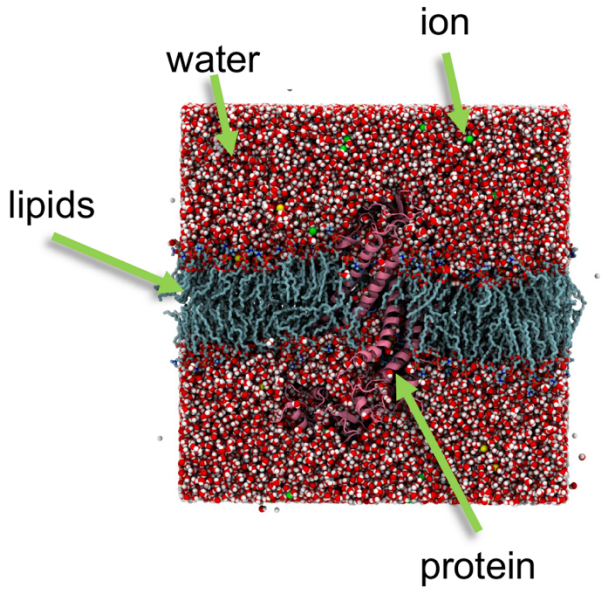

Figure 1.2 Snapshot from an AAMD simulation of TMEM16A embedded in a POPC lipid bilayer. TMEM16A (PDBI ID 7ZK3 [51]) is shown in pink, POPC lipids are shown as cyan sticks, water and ions shown as red/white, green and yellow VdW spheres.

The field has therefore invented variations on MD simulations that further simplify the system by mapping multiple atoms at a time to larger beads. This method, called coarse-grained MD (CGMD) simulation, however, still uses classical physics to solve for the velocities and positions of all beads in the system and has a customized forcefield that tries to match measurable features of fully atomistic simulations or experiments. In this style of simulation the conformations of proteins are more or less fixed by elastic restraints that prevent sampling of large conformational changes like those I observed in my atomistic simulations of TMEM16A. Due to the CG nature of the system representation and forcefield the energetic landscapes tend to be smoother and the kinetics are as much as 4x faster than in equivalent atomistic simulations and thus greatly expands the timescales one is able to achieve. I have successfully used CGMD simulations to sample hundreds of lipid scrambling events and dozens of ion permeations by different TMEM16 members which would not have been feasible with AAMD. Importantly, these simulations have been able to reproduce observations made at the atomistic level such as the shape of the membrane around TMEM16s and locations of scrambling events. To conclude, this thesis is a testament

to how both AA and CGMD can be used to investigate the molecular details underpinning protein functions, particularly the transport of ions and lipids across the membrane.

2. CHLORIDE CONDUCTION IN NOVEL STATES OF TMEM16A

Introduction

Amongst the characterized TMEM16 family members only two have been identified as true ion channels with no scramblase activity: TMEM16A and TMEM16B. TMEM16A was the first family member to be identified at a time when it was thought that TMEM16s were a family of Ca^{2+}-activated Chloride Channels (CaCCs) [9] [6] [7]. TMEM16A plays a role in a wide range of biological processes from gastrointestinal motility and fluid secretion to nociception and muscle contraction [73] [29]. It is also implicated in several cancers [31] and is a promising target to treat asthma and chronic obstructive pulmonary disease (COPD) [74]. In addition to Ca^{2+}, TMEM16A activity is also modulated by phosphatidylinositol-4,5-bisphosphate (PIP_2) which is in fact crucial for rescuing current following channel rundown under long exposure to high Ca^{2+} -concentrations [52]. It is however less clear if PIP_2 is necessary for TMEM16A activation [73][75].

Of the two TMEM16 ion channels only the structure of TMEM16A has been solved experimentally and in both apo and Ca^{2+}-bound states of the protein [63] [67][48]. Like other TMEM16 family members TMEM16A forms a homodimer with a butterfly-like fold and each subunit has a putative hydrophilic pore formed by TM3-7. However, none of the TMEM16A structures have pores wide enough to accommodate a bare Cl^- ion so it is unlikely that any of these structures reflect a conductive state of the protein. Of note, none of these structures were solved in the presence of PIP_2 (**Supplemental Information Table 2.1**). Hybrid continuum and AAMD simulations initiated from one Ca^{2+}-bound structure (PDB ID 5OYB [63]) were able to predict sites on TMEM16A that impact the ability of PIP_2 to rescue current and identified several possible lipid binding sites on the protein [52][49]. Later Jia *et al.* [50] ran long multi-μs AAMD simulations of the same Ca^{2+}-bound structure but with PIP_2 placed *in silico* into one of the predicted binding sites and observed spontaneous pore dilation events and even reported a single unbiased Cl^- permeation in the open pore for the first time. Their model for TMEM16A opening

relies on an allosteric network connecting interactions at the PIP_2 site formed by basic residues at the cytosolic termini of TM3-5 with pore dilation occurring on the opposite end of TM4.

Another important feature of TMEM16A activity is its voltage dependence. Under low (< 1 mM) Ca^{2+}-conditions TMEM16A has an outwardly rectifying current meaning it more easily carries a current at positive voltages [52][46]. On the other hand, under higher Ca^{2+} concentrations the current appears linear (Ohmic) and less selective amongst different anions [46]. It is generally believed that this change in ion conductance is related to both a Ca^{2+} and voltage-dependent conformational change. Interestingly, if you make alanine mutations to basic residues facing or near the pore (K588, K645, R515, and K603) you can restore the outward rectification under high [Ca^{2+}] [48] [46], but it is unclear if this is due to changes in the pore electrostatics or conformation of the pore.

In line with these electrophysiology experiments there is structural evidence for multiple Ca^{2+} -bound TMEM16A conformations. Nanodiscs structures of the mouse TMEM16A(a) splice variant (PDB IDs 6BGI and 6BGJ [67]) with either one Ca^{2+} or two Ca^{2+} atoms bound have a ~45° rotation of the extracellular end of TM3 compared to a cyro-EM structure of the mouse TMEM16A(ac) isoform. Later an inhibitor-bound cryo-EM structure of mTMEM16(ac) (PDB ID 7ZK3 [51]) also captured this rotation of TM3. Given that the IC50 (8.6-30.9 μM) of this inhibitor, 1PBC, negatively correlates with Ca^{2+} concentrations and voltage, it was hypothesized that it preferentially binds to the high-Ca^{2+} state of TMEM16A [70][46]. This implies that the state of TMEM16A with Ohmic currents and weaker ion selectivity, the same or very similar state inhibited by 1PBC, has this rotation in TM3. An important aspect of this rotation is that it moves positively charge residue R515, which faces toward the dimer interface on the apo state, directly into the pore pathway (**Figure 1.1**). This at least explains why R515 mutants have such a strong effect on TMEM16A selectivity [48] [46] as it likely comes into direct contact with ions in the pore. Moreover, reanalysis of the density from the earlier Ca^{2+}-bound structure (5OYB) indicates that both TM3 conformations are present in the density and therefore both states are likely visited at equilibrium.

To determine if the state of TMEM16A with the rotated TM3 indeed has Ohmic style current I have used atomistic MD simulations to explore the ion conduction properties of the 1PBC inhibited state, but once the inhibitor was removed. I initiated 12 simulations from the inhibited structure (7ZK3) and ran each for at least 1 μs. By the end of the simulations nearly half of the subunits spontaneously dilated their pores by moving apart residues forming a hydrophobic gate at the membrane midplane. These dilated states had notably increased density for chloride in the extracellular vestibule compared to a simulation with the pore backbone atoms harmonically restrained to the starting inhibited state. I observed six full chloride permeation events each proceeding after separation of the hydrophobic gate by more than 9 Å. The chloride ions appear to linger at three distinct locations in the pore and are coordinated by basic residues K588, K645 and R515. In addition to opening the pore some of the remaining subunits further constricted the pathway. Open and closed states appear to bend or "kink" TM4 at two different locations. I then clustered the aggregate simulation data to help identify the major conformations sampled by the inhibitor-released TMEM16A and initiated new simulations from multiple states to see if any of them conducted ions under an applied voltage. Two of the states with pore radii greater than the ionic radius of chloride indeed did sample multiple permeation events, up to 20 for a single subunit. Unexpectedly, the most dilated state (cluster 12) had more than double the number of events at 300 mV compared to -300 mV. This difference appears to be coupled to the rotamer state of K645 which points outward under positive potentials and inward under negative. The inward-facing conformation forms a favorable Cl^- binding site in the cytosolic vestibule which slows the rate of transport. However, the electronic behavior of the dilated states is complicated by the fact that lipid headgroups can also interact with residues in the outer pore entrance and temporarily block ion permeation.

Results

Simulations reveal a dilated ion pore upon 1PBC removal

The first question I wanted to address was whether the 1PBC-bound mTMEM16A conformation captured by cryo-EM would remain stable in a simulated lipid bilayer in the absence of 1PBC. In other

words, could I see with simulations if 1PBC is required to maintain this conformation with a widened extracellular vestibule. However, before initiating any simulations, I needed to address several cytosolic loops that are unresolved in the 1PBC/Ca^{2+} -bound cryo-EM structure (PDB ID 7ZK3 [70], **Figure 2.1*A***). The loops preceding TM3 and TM7 are often missing in TMEM16 experimental structures likely due to their inherent flexibility. I was particularly concerned with the placement of the pre-TM3 loop as residues there have been implicated in an allosteric network for pore opening and could affect the behavior of both TM3 and 4 (which form half of the ion pore) in the simulations [50]. Given how long two of the loops are (15 and 13 residues) I decided to compare models of the residues from two different structure prediction algorithms: MODELLER [76] and AlphaFold [77] (see Methods for details) (**Supplemental Information Figure 2.1**) to find the most reasonable starting conformation. In the end, the positions of these loops varied so greatly from one another that I decided to simulate both sets of predictions. I generated three different models, 2 symmetric and 1 asymmetric, with loops from either method as well as a fourth model with none of the cytosolic loops included as a control.

To test the stability of the 1PBC-inhibited state I removed the inhibitor coordinates from each TMEM16A partially modelled structure prior to minimization. I performed three replicates of all-atom MD (AAMD) simulations of each model for 1 μs which resulted in an aggregate simulation time of 24 μs when accounting for both dimer subunits. Surprisingly, I observed 10/18 subunits with modelled cytosolic loops and 3/6 subunits without spontaneously undergo major conformational changes that increased the distance between residues on TM4 and 6. These changes were most evident at the central, initially most constricted, part of the pore formed by the hydrophobic gate [78] (**Figure 2.1*B-C***). Each of these events occurred with the first 300 ns of each simulation indicating the starting structure was unstable and as a result quickly relaxed to these new states. To quantify this pore dilation, I measured the minimum distance between two residues part of the hydrophobic gate: L547 (TM4) and I641 (TM6) over the course of each simulation. During opening events the minimum distance between these residues increases up to ~6 fold the starting cryo-EM value. More interestingly, the distribution of these distances from my

aggregate simulation data reveals a multimodal landscape with peaks around 4.5-5, 7-8, and 10-11.5 Å indicating that our simulations may have been sampling multiple metastable conformations (**Figure 2.1*D***). The presence of multiple Ca^{2+}-bound states, corroborates results from noise analysis of macroscopic TMEM16A currents indicating that the protein naturally samples "pre-open" Ca^{2+} -bound intermediates [79]. The smallest dilation (dilated 1 in **Figure 2.1*D***) appears to be characterized by a slight rotation and concomitant bend of TM4 (as seen in **Figure 2.1*A***) at L547 causing the residue sidechain to move out of the pore interior. The wider dilations (dilated 2 and 3) seem to involve larger-scale motions of TM4 including tilting into the membrane. In addition to seeing the pore dilate we also saw it constrict which is represented by L547-I641 distances around 2.2 Å (~3.4 Å when excluding hydrogen), closer than the starting cryo-EM coordinates. In fact, this distance is similar to values captured by apo TMEM16A structures (**Supplemental Information Table 2.1**). However, in all simulations, even those with constricted states, all Ca^{2+} ions remain bound (data not shown). The sampling of these distances also appears to vary between protein models with different cytosolic loop predictions. For example, simulations with symmetric MODELLER-based loops have increased relative sampling around the 4.5-5 Å range, while the asymmetric MODELLER-based loop model spends little time sampling these distances (**Supplemental Information Figure 2.2**). Contrastingly, asymmetric MODELLER-based and unmodelled loop models both sample the constricted distance more frequently than other models.

I next looked at how these pore dilation changes altered ion accessibility to the pore. I measured the total density of either chloride or potassium ions along a 3D path coordinate centered on the pore (see **Methods** and **Supplemental Information Figure 2.3**) from aggregate simulation data with L547 and I641 distances 0-3 Å (constricted), 3-6 Å (dilated 1), 6-9 Å (dilated 2), or 9-20 Å (dilated 3). In wider conformations (dilated 1-3) there is a clear increase in the density of chloride throughout the pore with notable peaks below and above the hydrophobic gating residue L547 (**Figure 2.1*E***). Constricted conformations have a similar peak below L547 (intracellular vestibule), although with weaker density. I also ran a 1 μs simulation initiated with asymmetric MODELLER-based loops with 1 kcal/mol·A^2

backbone restraints to assess the ion accessibility of the inhibited state. Like the predicted constricted states, the density of Cl^- was near zero above L547 (extracellular vestibule), but the restrained simulation had a much higher peak in the cytosolic vestibule, even higher than the dilated states. In dilated, constricted and inhibited conformations the density of potassium is nearly zero in the pore which is consistent with experimental evidence that TMEM16A is anion selective (**Figure 2.1*F***) [47].

To test if these observed conformational changes were due to the removal of 1PBC I also initiated simulations with the inhibitor still bound. Experimental observations that 1PBC preferentially binds at depolarizing voltages and the anion selective nature of TMEM16A imply that 1PBC binds in a deprotonated, negatively charged state, despite the high predicated pKa (9.39) of the hydroxyl in solution [51]. The pose of the molecule in the pore also indicates that the negatively charged 1PBC oxygen may be interacting with nearby K603 (TM5) and R515 (TM3) which when mutated significantly reduce the inhibitor's potency [51]. However, others have shown that these two residues are involved in TMEM16A's current voltage-dependence which muddies their potentially direct effect on inhibitor binding [48] [46]. I therefore decided to simulate both the anionic and neutral states of the inhibitor initiated from the same cryo-EM pose and performed three replicates of each for 1 μs. Both the charged (1PBC(-1)) and neutral (1PBC(0)) forms of the inhibitor stay bound to the pore in all simulations. However, I did observed two instances in separate simulations when 1PBC(-1) moved deeper into the pore away from K603 but continued interacting with R515 which tilted its sidechain down toward the center of the pore to meet the descended inhibitor (**Supplemental Information Figure 2.4*B, right***). Otherwise the 1PBC(-1) oxygen atom maintains distances to K603, R515 and E633 similar to the cryo-EM pose (**Supplemental Information Figure 2.4*D***). Meanwhile the 1PBC(0) protonated oxygen shifts away (~ 2 Å) from K603 and closer to E633 (**Supplemental Information Figure 2.4*C,E***). Additionally, the extracellular vestibule moves up to 4 Å away the cryo-EM position when either 1PBC form is bound (**Supplemental Information Figure 2.4*D,E***), however this is still lower than some dilated states when 1PBC is removed (**Supplemental Information Figure 2.5**). When 1PBC(-1) is bound TM4 bends and

separates L547 and I641 to the same degree as seen with 1PBC removed (**Figure 2.1*D***). When 1PBC(0) is bound the gate also dilates but not as widely as some instances with 1PBC(-1) and the upper half of TM4 (above L547) tilts away from the pore **Supplemental Information Figure 2.4*C***. Despite these changes to the binding pose, I did not observe any Cl^- permeation events during these simulations or even any ions in the extracellular vestibule. Taken together my results show that 1PBC acts as an effective pore blocker in either its anionic or neutral state, but that the protein still moves away from the cryo-EM pose in a simulated lipid bilayer, although how it does this seems dependent on the 1PBC protonation state. 1PBC is also a weak binder (~3.6 μM K_d at 0 mV) so it not unexpected that the inhibitor moves around in the binding pocket during simulations [51].

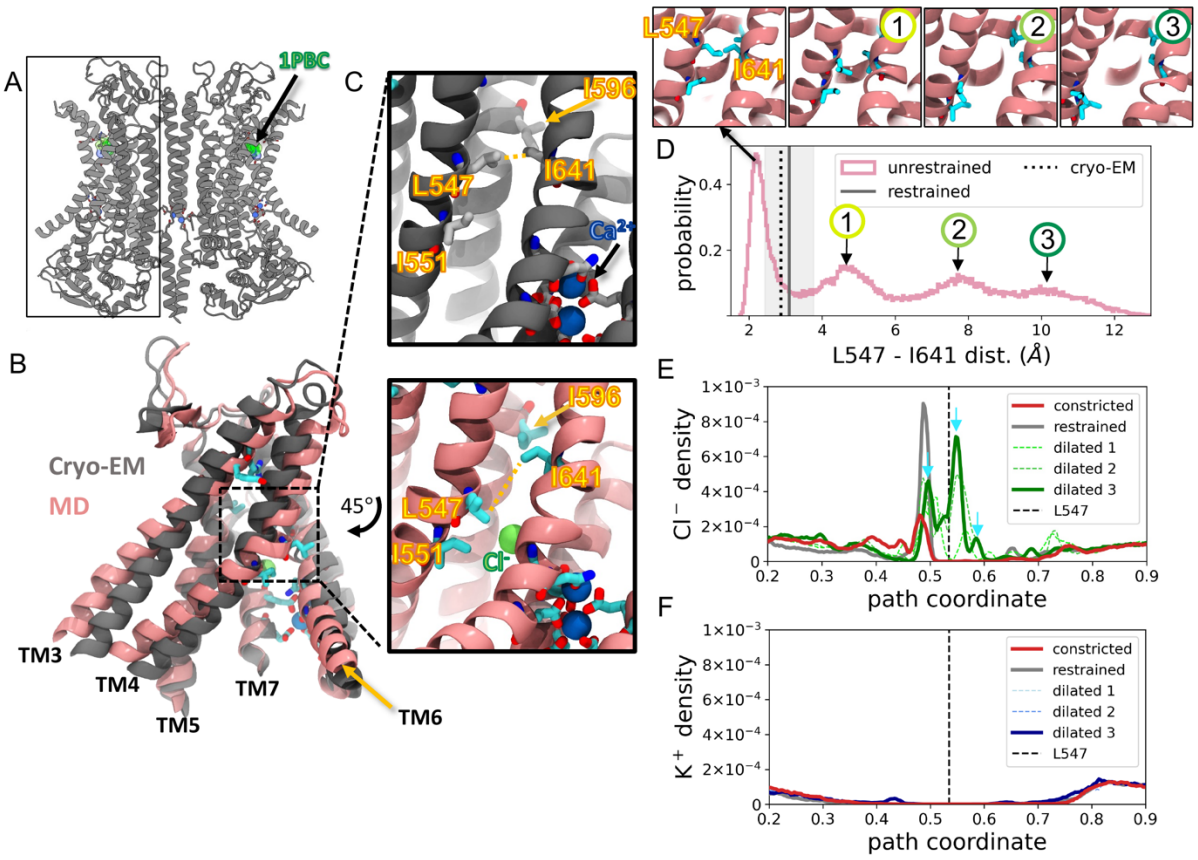

Figure 2.1 TMEM16A pore constricts or dilates the ion pathway when 1PBC inhibitor is removed. A) Cryo-EM structure of 1PBC/Ca^{2+} -bound TMEM16A (PDB ID 7ZK3). B) Overlay of the starting cyro-EM TMEM16A coordinates (grey) and snapshot from an atomistic MD simulation. (*Figure caption continued on the next page.*)

(*Figure caption continued from the previous page.*) C) Zoomed-in view of pathway hydrophobic gate in the cryo-EM (top) and same simulation snapshot in (B). The chloride ion is shown as a green sphere. D) Histogram of the minimum distances between L547 and I641 from aggregate simulation data with simulation snapshots from each peak in the distribution (above). Starting distance (cryo-EM) is indicated by black dotted line. Standard deviation and mean of distances from a 1 kcal/mol·A^2 backbone-restrained simulation represented by the grey line and shaded region. E and F) Average density of chloride and potassium along the pore from simulation data with L547-I641 distances 0-3 Å (constricted), 3-6 Å (dilated 1), 6-9 Å (dilated 2), and 9-20 Å (dilated 3). The density from a 1 kcal/mol·A^2 backbone-restrained simulation represented in grey. The pore pathway is calculated from a weighted average of the water density from all simulation data (see SI Fig. 2.3 and Methods) and density collected with 6 Å of the pathway center. Distance and density analysis only includes simulation data collected after 500 ns. Density values are averaged across data collected from A and B subunits. A rolling window average has also been applied to the density values every 2.5 Å.

Dilated states of TMEM16A are ion conductive

In addition to seeing the increase in Cl^- density when the pore in dilated states I observed 6 full ion permeation events in independent simulations with 1PBC removed: 5 from the cytosolic bath ('upward') and 1 from the extracellular bath ('downward') (**Figure 2.2*A***, **Supplemental Information Figure 2.6**). Each of these permeation events occurred when the L547-I641 distance was at least 9 Å. It is evident from traces of the ion positions in the pore that they tend to spend more time at three particular z-values (-10 Å, 0Å, and 10 Å) during their permeation events (**Figure 2.2*A***) which corroborates the three peaks in the Cl^- density profile (**Figure 2.1*E***). Here on I refer to these locations as sites A, B and C. When I closely inspected these locations in the simulations I saw that site A, which is the least occupied according to the density profile, is the same as the 1PBC binding site (**Figure 2.2*B, top***). Indeed Cl^- ions are simultaneously coordinated by R515 and K603 at the top of the extracellular vestibule like the hydroxyl oxygen in 1PBC(-1). Site B is the second most occupied location and is formed by K645 (TM6) and K588 (TM5) in the cytosolic vestibule (**Figure 2.2*B, bottom***). Lastly, Site C, which sits between the A and B sites just above the former pore constriction site is formed by N546, K645 and R515 (Figure 2.2*C*). The latter two residues dynamically take on different rotamer states to coordinate Cl^- in the middle. We observed that K645 (also noted by Jia *et al.* [50]) and R515 would snorkel up and down to maintain

contact with ions in the pore depending on their position. This means that at least two of the Cl^- binding sites in TMEM16A are incredibly fluid.

As the ion moves through the pore it is typically accompanied by at least 3 close water molecules and interacts directly with polar and negatively charged residues. K645, K588, and R515 spend 10-34% of the aggregate simulation time in contact with at least one Cl^- ion (**Figure 2.2*D)***. Although most events last less than 200 ns there are a few examples of interactions lasting over 400 ns, nearly half of the individual simulation times. Comparably K603 (site A), R535 above site A, and R562 (TM4) near the intracellular pore entrance also contact ions for 10s of ns but are ~4-8 fold less frequently in contact. Other basic residues near the pore that occasionally contact Cl- (K741 (TM6), K327 (TM1), and K661 (TM6)) are considerably farther from the high ion density locations (**Supplemental Information Figure 2.7**).

Alternative Kink Sites on TM4 control whether the pore is open or closed

Once I established that these dilated states could permeate ions I wanted to better understand why around half (13/24) of the simulations moved to open states (dilated 3), 7/24 closed and remaining 4 have some intermediate dilation (dilated 1-2). I already noted that simulations with open structures kink TM4 at L547 which also causes a secondary structural change to a π-helix bulge at the same location which starts to separate the hydrophobic gate (**Supplemental Information Figure 2.8*A)***. Later I noticed that closed structures also kink TM4 but at a different location, approximately two helix turns lower around E555 (**Supplemental Information Figure 2.8*B)***. This second kink also seems to force the extracellular terminus of TM4 further into the pore. I never observed bending at both sites concurrently, but there are a few examples where neither location forms a kink, and the protein resembles the starting structure. Surprisingly, neither of these locations contains a glycine or proline residue which tend to break helical structure.

I observed clear changes in the number of waters in the center of the pore as a function of the helix bend angle at L547 or E555. As the L547 kink angle decreases (15-20°) the number of waters increase

and as E555 angle decrease (30-40°) the number of waters drops to nearly zero essentially dehydrating the pore (**Supplemental Information Figure 2.8*A-B right,* Supplemental Information Figure 2.8*D***). There is also a strong positive and negative association between the L547-I641 distance and the kink angle at L547 and E555 respectively (**Supplemental Information Figure 2.8*D***). When I applied restraints to maintain the inhibited conformation, which has a relatively straight TM4 with angles at both locations ~175°, there was an intermediate amount of water in the pore compared to the open and closed states (**Supplemental Information Figure 2.8*C*)**. The presence of these kinks indicates that the starting conformation of TM4 comes under some kind of pressure when the protein is put into a simulated bilayer and as a result buckles during simulations. Moreover, including 1PBC or models of cytosolic loops does not prevent TM4 from moving away from the starting structure, and thus the deformation in our simulated POPC bilayer is likely driven by some other structural feature of the inhibited state. It is still unclear why the protein moves to one state versus the other.

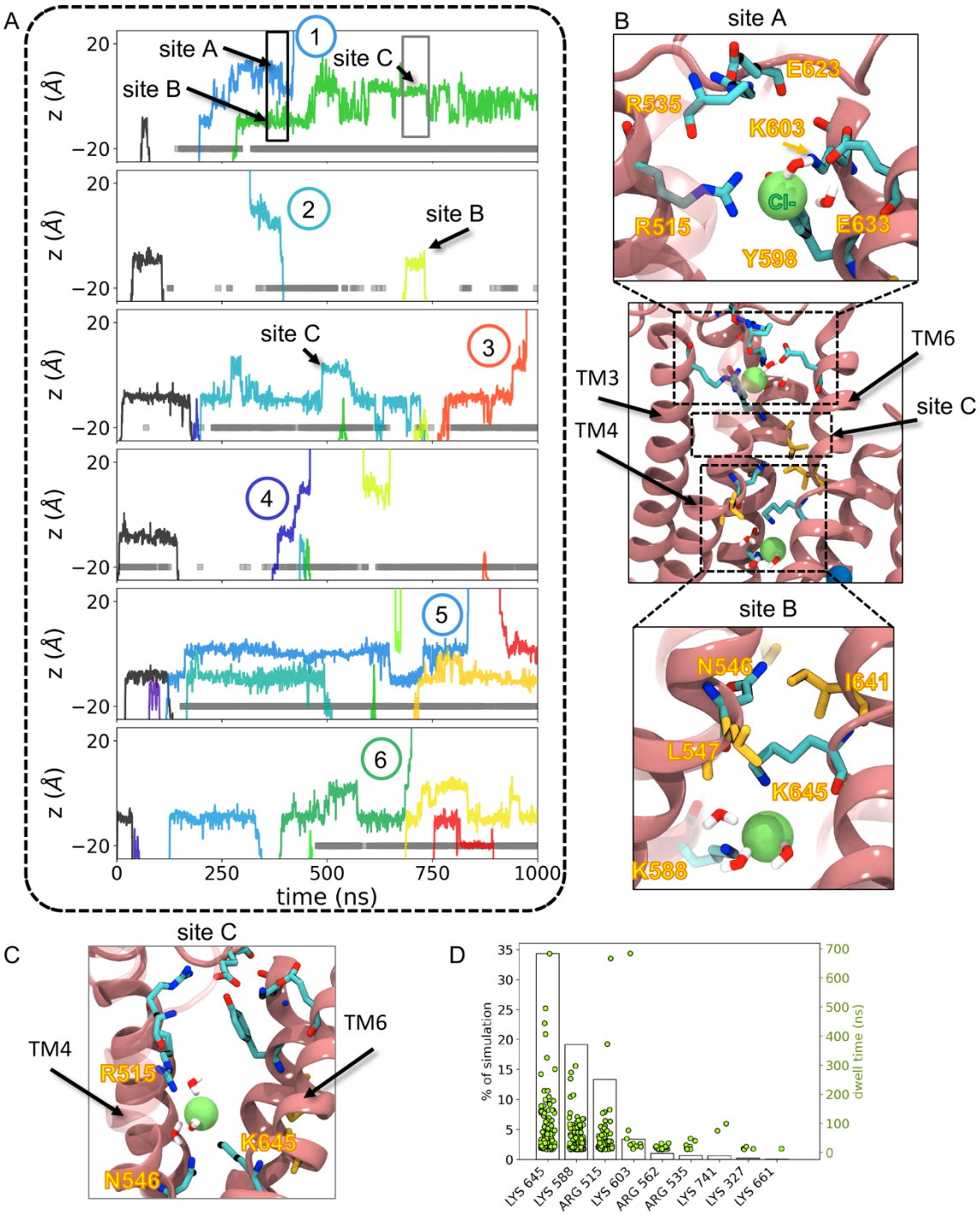

Figure 2.2 Spontaneous chloride permeation through dilated states of TMEM16A.
A) Z-positions of chloride ions in contact with the pathway, each color represents a unique ion. The y-axis is zeroes at the cryo-EM z-position of the L547 Cα. Only simulations containing complete permeation events (numbered) shown here. The grey bars indicate when the L547-I641 distance is > 9 Å. B) Snapshots of the open TMEM16A groove with zoomed-in images of two Cl^- interaction sites: site A (top) and site B (bottom). (*Figure caption continued on the next page.*)

(*Figure caption continued from the previous page.*) C) Zoomed-in snapshot of the third Cl^- interaction site (site C).D) Plot of the percentage of simulation time basic residues interact with Cl^- ions (left y-axis, black bars) and dwell times of each interaction instance (right y-axis, green dots).

TMEM16A conformation changes involve a small hydrophobic network between TM3 and TM4

Up until this point I have established that dilated states the states of TM4 can conduct ions, but it is unclear which type of currents these conformations convey: Ohmic or outwardly rectifying. Additionally, all three dilated states had some density for Cl^- in the extracellular vestibule (**Figure 2.1*E***) so it could still be possible that ion conduction also occurs in the less open states, but that they have not had sufficient time to sample it at 0 mV. To help distinguish distinct conformations in my systems I applied the dimensionality reduction method time-independent component analysis (tICA) [80] on a set of 20 residue-to-residue minimum distances between TM3, 4, 5, and 6 (**Supplemental Information Figure 2.9A, B**). I then performed MiniBatch K-means clustering on the first 13 independent component (tIC) dimensions and constructed a Markov State Model (MSD) using the 500 microstates from clustering. Finally, I selected 15 clusters (i.e. macrostates) using the Robust Perron Cluster Analysis (PCCA++) [81] [82] which provide sufficient coverage of the simulation data spread over at least the first four tICs (**Supplemental Information Figure 2.9C-F**).

Most of the simulation data in the projection onto the first two components (tICs) is concentrated near the starting cryo-EM structure and spreads out in two principal directions (**Figure 2.3*A***). After carefully inspecting conformations associated with each cluster I concluded that moving to the left of the starting structure (decreasing values of tIC 1) is related to groove dilation (i.e. increases in water accessibility) that results from widening L547-I641 distances, decreases in the TM4 kink angle at L547 and tilt of TM4 (**Figure 2.3*B*, Supplemental Information Figure 2.10** and **2.11*B,E,F***). Meanwhile, increasing values of tIC 2 is qualitatively correlated with increasing distances between the intracellular ends of TM3 and 4 which doesn't have a significant impact on how open the pore is (**Supplemental Information Figure 2.10** and **2.11*B, F***). I also observed a striking change in the secondary structure of

the extracellular end of TM3 as the groove dilated. Upon close inspection there is a small network of hydrophobic residues (L522 on TM3 and I534/V538 on TM4) that become separated in open states farther from the cryo-EM structure in the tIC space like clusters 12 and 14 (**Figure 2.3*C, upper panel***). This network is not present in structures without the rotation of TM3. I also noticed that even though TM4 underwent these large conformational changes as the pore dilated it still maintained contact with TM6 although with a different set of interactions. As TM4 bends and tilts it does so toward TM6 which causes both I551 and L547 (TM4) to encounter F653 (TM6) (**Figure 2.3*C, lower panel***). Maintaining contact between TM4 and 6 could be very important for preventing lipids to scramble in the pore as it dilates which occurs in TMEM16 members that scramble. F653A mutations do not significantly impact TMEM16A activation [47], but it would be interesting to explore if it conveys scramblase activity.

Consistent with dissimilarities in the L547-I641 distance distributions (**Supplemental Information Figure 2.3**) I observed stark differences in how the tIC space was sampled by each TMEM16A model (**Supplemental Information Figure 2.12*B***). All of the data that contributes to clusters 4-6 comes from simulations without cytosolic loop predictions, which also fail to explore the lower half of the tIC 1 space. The models with either AlphaFold-based and chain B of MODELLER-based asymmetric loop predictions also have a limited range in tIC 1 and tend to stay close to the cryo-EM position. Chain A of simulations initiated from MODELLER-based asymmetric loop predictions and the fully symmetric model are the only simulations to sample the extreme end of open states, clusters 11-14 (**Supplemental Information Figure 2.12*B,C)***. I speculate that these differences may largely be driven by either the position or absence of the pre-TM3 loop which varies between the models and is proximal to TM3 and 4 which move the most during simulations.

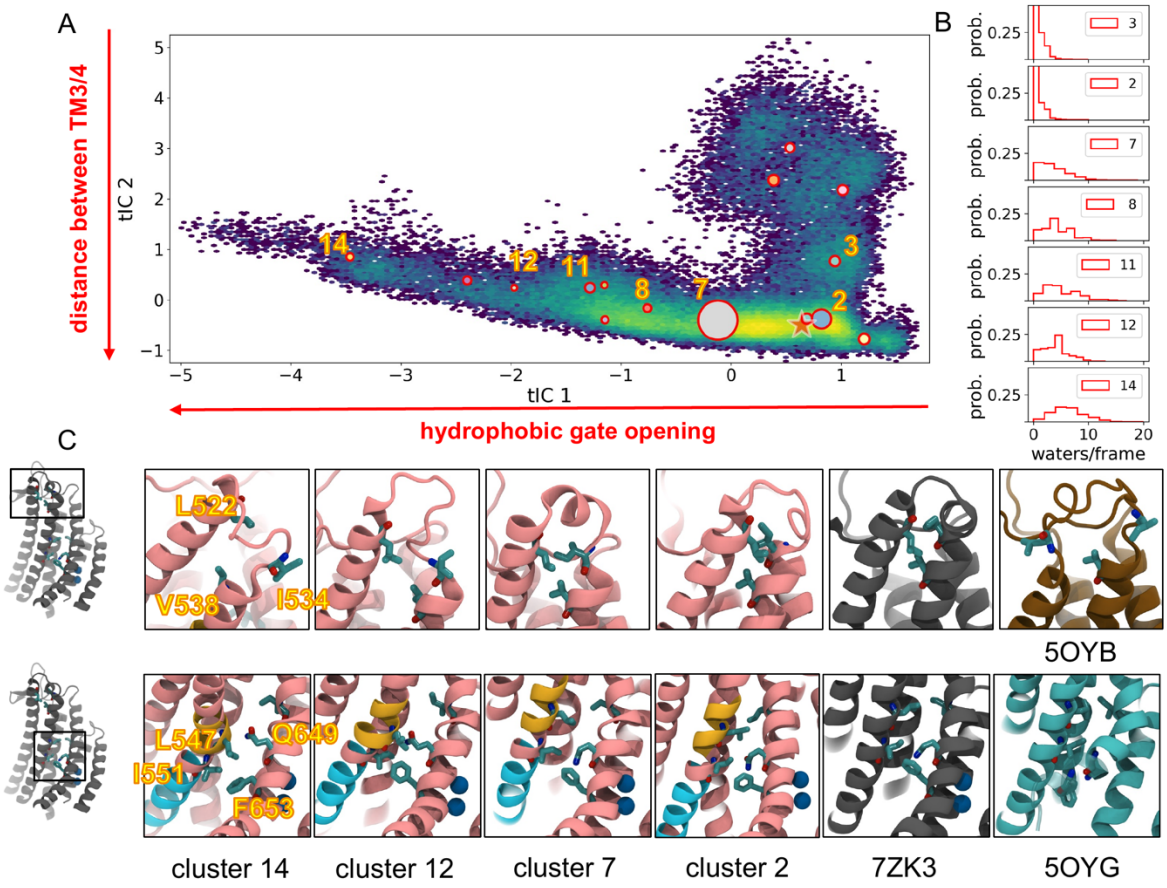

Figure 2.3 TMEM16A pore opening involved rearrangement of hydrophobic contacts.
A) All simulation data projected onto the first two tIC eigenvectors colored by log density (13 tICs used in clustering, see Methods for details on tICA) . Circles indicate macrostate (cluster) centers sized by relative population. (B) Histograms of water flux (number of waters/frame) through the center of the pore in select cluster. C) Medoids of selected clusters. Top: hydrophobic network at TM3/4 extracellular termini. Bottom: hydrophobic contacts between TM4 and 6. Cryo-EM structures of 1PBC/Ca^{2+} -bound, Ca^{2+} -bound and apo TMEM16A shown in grey, brown and cyan respectively.

Simulated TMEM16A currents appear linear despite voltage-dependent change in basic residue position on TM6

To quantify the conduction properties of the various states captured by my simulations I then wanted to simulate states with an applied voltage across the membrane. I selected 5 clusters from the tIC and clustering analysis (clusters 2, 8, 7, 11 and 12) and initiated new simulations from the medoid structure of each cluster. I choose these states because they were highly represented in the aggregate data

(cluster 2 and 7 are the largest clusters), overlapped with sampling from all simulations (except for AlphaFold-based loop models) and provided a representation of a variety of structural features. This includes disengagement of the TM3-4 hydrophobic network (cluster 12) and severe kinking of TM4 (cluster 8). The average pore radii calculated for each cluster indicates that only one of the clusters I selected has a pore wide enough to accommodate a bare Cl^- ion (cluster 12) (**Fig 2.4*A***). However, clusters 7 and 11 minimum pore radii are within 0.1 Å of the Cl^- ionic radius and perhaps local sidechain rearrangements at the constriction site could let the anion through. I simulated each structure for 1 μs each at -300, -250, -200, -150, 0, 150, 200, 250, and 300 mV and applied 1 kcal/mol·A^2 restraints to the backbone heavy atoms of the pore-delineating TM3-8 to maintain the starting conformation.

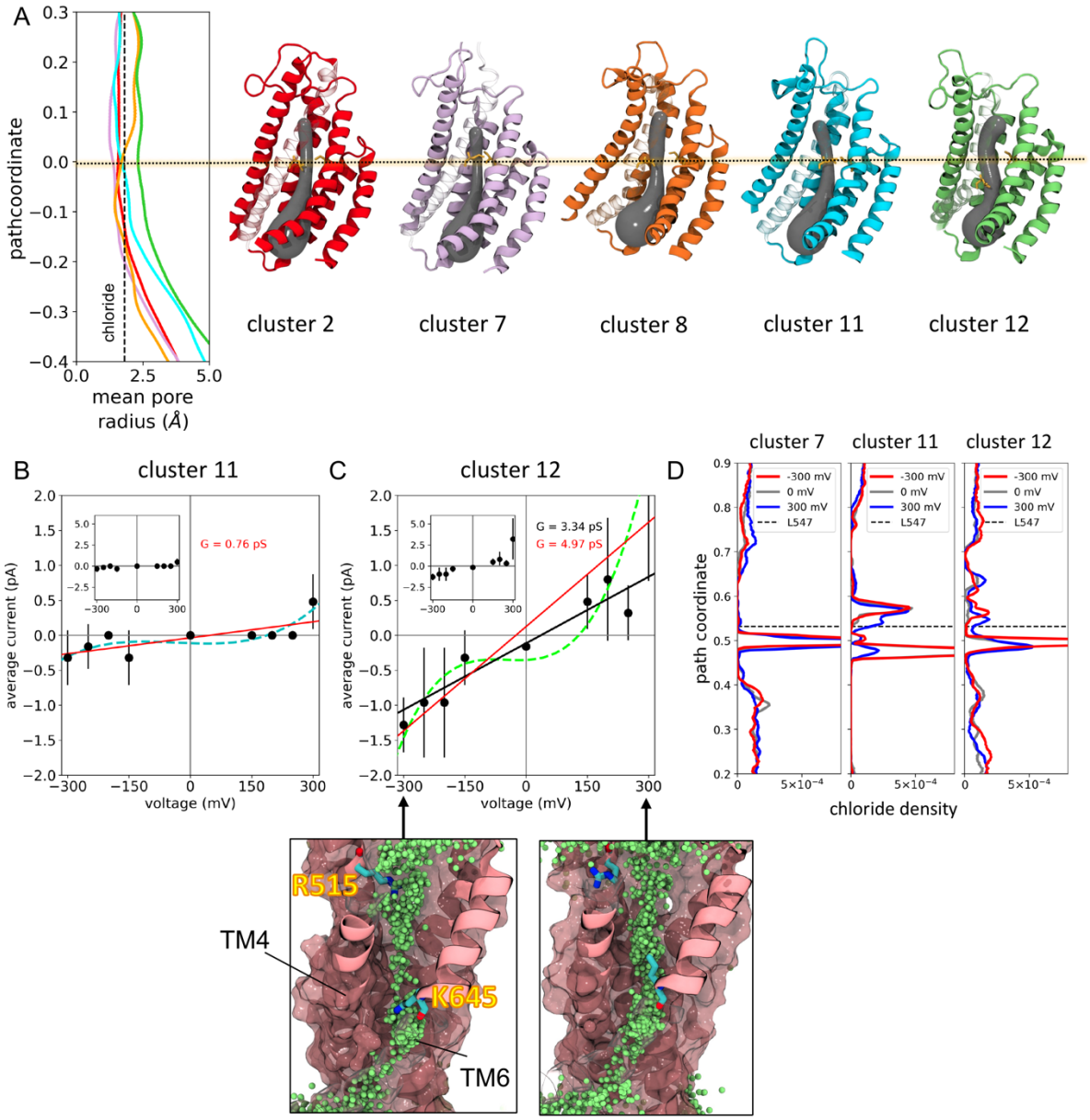

Figure 2.4 Dilated TMEM16A state has outward-rectifying-like current.
A) Average pore radius of TMEM16A clusters 2 (red), 7 (purple), 8 (orange), 11 (cyan), and 12 (green) (see methods for pore size calculations with HOLE2). Medoids of each cluster with 3D spline (grey) fit to averaged pore center coordinates and radii from the same cluster (right). The pore path coordinate is zeroed at the average K588 (TM5) and K645 (TM6) sidechain nitrogen z-positions. L547 (TM4) and I641 (TM6) shown as yellow sticks. Dotted yellow line represent the pore midpoint on each state. B, C) Currents calculated from simulations of cluster medoids plotted as a function of voltage. 1 kcal/mol·A^2 positions restraints were applied to the pore-lining helices of each applied voltage simulation. Curves were fit using a cubic (dotted line) and linear (solid red line) 1D polynomial and error bars indicate the STDV of currents. Liner fit for data points -250 to 250 mV show as black line in (C). Inset conductance values shown in corresponding colors. (*Figure caption continued on the next page.*)

(*Figure caption continued from the previous page.*) Currents and there STDV were taken from averages over 200 ns simulation windows. (C, below) Cartoon representation of the cluster 12 medoid during simulations under an applied -300 mV (left) and 300 mV voltage with overlay of all Cl- positions every 0.5 ns. D) Chloride density along the pore pathway for cluster 7, 11 and 12 at -300, 0, and 200 mV. The pathway was calculated from a weighted average of the water density from all simulation data (see **Methods**). Starting position of the L547 Cα (cryo-EM) relative to the pore patheway is indicated by black dotted line.

Consistent with the pore radii predictions I did not observe any ion permeation events for cluster 2 which earlier we referred to as a constricted state. I also did not observe any full permeation events for cluster 8 and we only saw two permeation events for cluster 7: one at +300 and another at +200 mV. In total we observed 10 permeation events for cluster 11 and 54 events for cluster 12, our two open states (**Table 2.1**). In each case the ion traveled with its electrical gradient. I also did not observe any potassium permeation events. Average currents for both the cluster 11 and cluster 12 sets of simulations are in the pA range (**Figure 2.4*B, C***) but there appears to be an asymmetry in the number of events at positive and negative voltages for cluster 12 (**Figure 2.4*C***). There were more than double the number of events at 300 mV for the cluster 12 than at -300 mV and accordingly the maximum conductance (among 200 ns blocks of simulate) was ~4x larger at 300 mV than -300 mV. However, excluding this spike in events at 300 mV, the current-voltage relationship looks linear or Ohmic for both cluster 11 and 12 and consistent with the experimentally observed linear current seen for TMEM16A associated with saturating [Ca^{2+}] [46][47]. The conductance values (slope of the linear fit to the mean currents) for both clusters (0.76 and 3.34-4.97 pS) are also within the range from experiments (0.5-8 pS in experiments [83]). Although the maximum conductance value within a 200 ns window at 300 mv is more than double that of experiments and will be discussed later (**Table 2.1**). I also observed a slight change in the distribution of chloride ions in the pore when comparing simulations at -300 and 300 mV for clusters 11 and 12 in that at positive voltages the positions of ions appear more continuous through the middle of the pore (**Figure 2.4*C***, **Supplemental Information Figure 2.13*A-B***). However, the clearest change in the ion density profiles for the same clusters and voltages is the decreased density at site 2 (below L547) at 300 mV (**Figure 2.4*D***). On the

other hand, there doesn't appear to be any voltage-dependent changes in the ion density profile in simulations initiated from cluster 7.

Table 2.1 Number of chloride permeations events and maximum conductance in simulations of cluster medoids with applied voltage.

	Number of Events			G (pS)		
Voltage (mV)	7	11	12	7	11	12
-300	0	2	8	0	2.67	5.34
-250	0	1	7	0	3.20	6.41
-200	0	0	7	0	0	12.02
-150	0	2	2	0	5.34	5.34
0	0	0	1	-	-	-
+150	0	0	3	0	0	5.34
+200	1	0	6	3.20	0	12.01
+250	0	0	2	0	0	3.20
+300	1	3	20	5.34	2.67	21.36

At least in simulations initiated from cluster 12 I noticed a clear difference in the position of K645 at positive and negative voltages. At 300 mV the sidechain tends to be tilted outward toward site A (forming site C) and at negative -300 mV it tends to point inward to form site B (**Figure 2.5*B, bottom***). I already noted that in the absence of a membrane potential this residue side chain moves up and down to coordinates passing ions. R515 also appears more dynamic than other charged residues in site A and seems to rotate in the XY plane between pointing toward the pore center or into the TM3/4 interface (**Figure 2.5*B, top***). I measured both the rotation of K645 (with respect to the z-axis) and R515 (with respect to the x-axis) for all simulations initiated from MSM clusters and each applied voltage (**Figure 2.5*C***). Although the mean angle of R515 changes at some voltages for cluster 12, there is not a clear connection to the strength and sign of the applied membrane potential. Anecdotally, I've seen R515 interact with lipid headgroups through the TM3/4 interface for long periods (100s of ns) that seem to hold it there, but this doesn't happen in every simulation. The rotation angle of K645, on the other hand,

clearly changes in a voltage dependent manner with larger angles (sidechain pointing outward) consistently sampled at positive voltages and inward deflections at negative. This trend is most obvious for cluster 12 but is also seen in simulations of cluster 11 and not all present for the more constricted pores. Thus, the pore must be sufficiently dilated (L547-I641 distances at least > 7 Å) to allow K645 to flip up and down. Most interestingly, in cluster 12 at +300 mV there is also an increase in the percentage of simulation time K645 and R515 spend in contact with chloride and there are more K645 interaction events with dwell times >100 ns (**Supplemental Information Figure 2.15**). However, this voltage-dependent manipulation of ion permeation kinetics does not result is strongly rectifying behavior of the current.

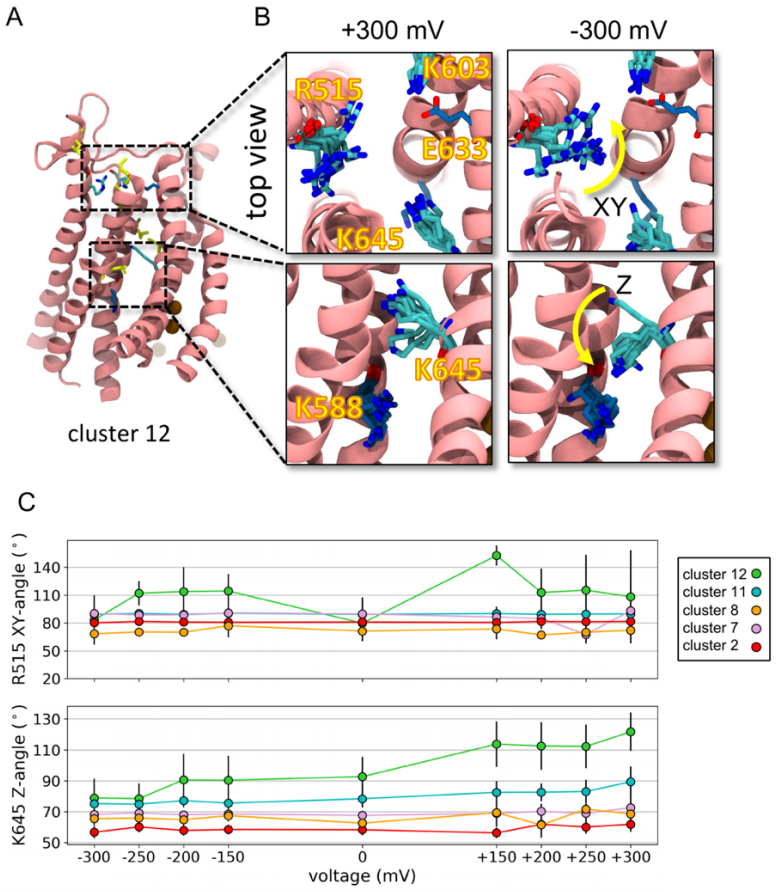

Figure 2.5 Position of basic residue lining the TMEM16A pore is voltage dependent.
A) Snapshot of a dilated macrostate (cluster 12) representative structure from a simulation with an applied positive voltage. (*Figure caption continued on the next page.*)

(*Figure caption continued from the previous page.*) B) Zoom-in on site A and site B from simulations with + or – 300 mV applied voltage. Cartoon representation of residue positions at multiple evenly spaced time points. C, top) The angle between the R515 principal axis vector and vector (1,1,0) (rotation in the xy-plane) and C, bottom) between K645 principal axis vector and the z-axis (rotation in xz-plane) for simulations of cluster medoids with applied voltage. Simulations were run with 1 kcal/mol·Å backbone positional restraints on pore-lining helices.

Lipids block ion permeation in most dilated state of TMEM16A

Even though the average current in simulations of TMEM16A initiated from cluster 12 is the greatest at 300 mV, so is the standard deviation of the mean. In the traces of chloride z-positions in the pore at 300 mV there is a consistent flow of ions inward between 50 ns and 500 ns, but then it abruptly stops and only one permeation event is seen from then on to the end of the simulation (**Figure 2.6*A***). I observed the same phenomenon at 200 mV. This explains why the maximum conductance in 200 ns windows was so much higher than the value fit to the mean currents (**Figure 2.6*C,* Table 2.1**). When I inspected these two simulations, I saw that this drop in current was associated with a POPC lipid headgroup inserting into the extracellular entrance of the ion pore and having its choline groups interact with E623 on TM5 (**Figure 2.6*B***). This also occurs at negative membrane potentials (**Figure 2.6*C***). Lipids binding into the pore however do not seem to completely block ions from getting through as there are several examples of ions getting through even when lipids are in the pore (grey bars in **Figure 2.6*A***). I did not observe lipid block or partial block of chloride in simulations of cluster 11, as this conformation seems to exclude lipid headgroups from the pore (**Figure 2.6*D, E***). Of all of the simulations initiated from MSM clusters, cluster 8 which has the most sever bend in TM4 at L547 has the highest density of lipids in the extracellular vestibule (**Supplemental Information Figure 2.14*B***).

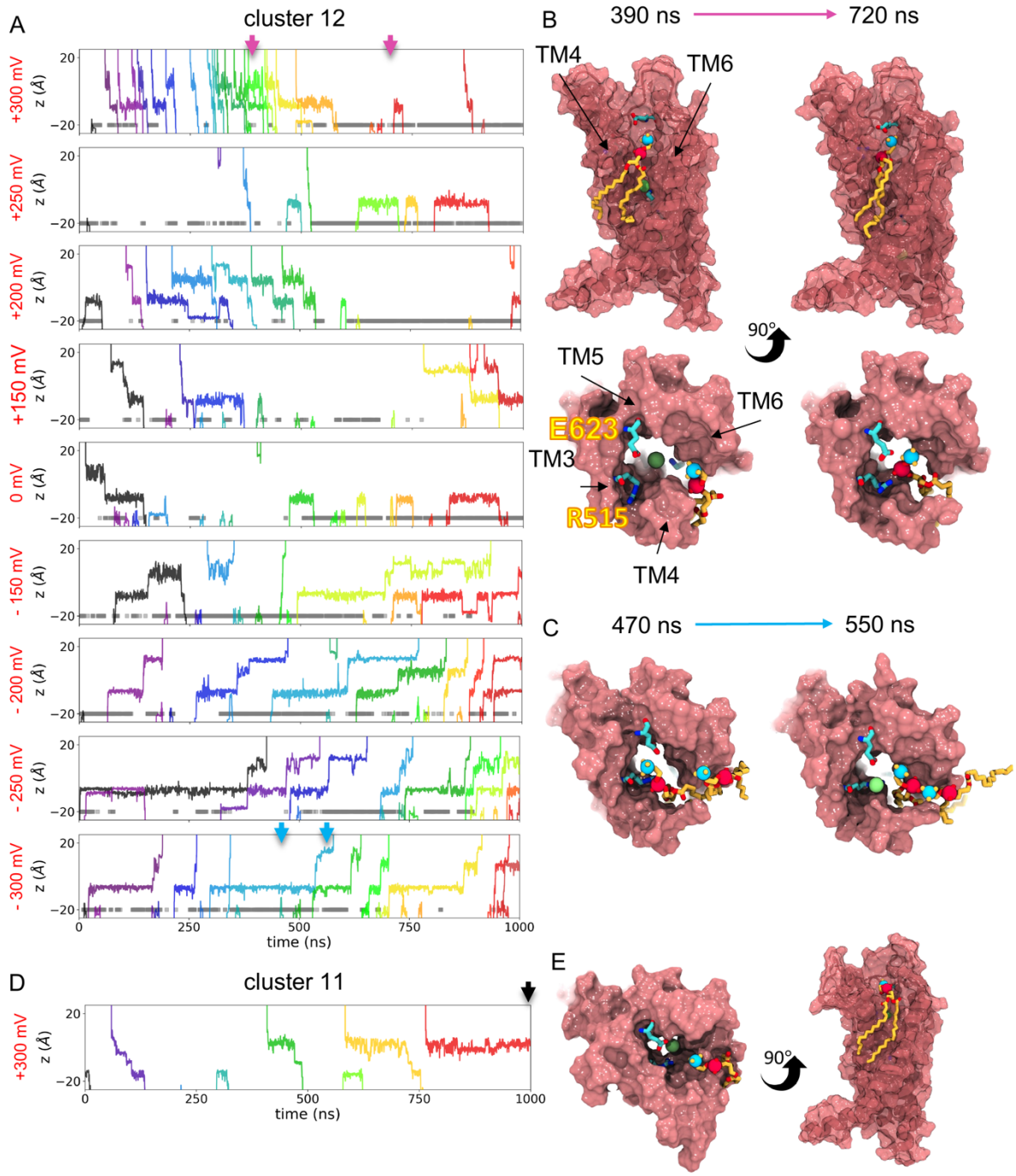

Figure 2.6 Lipids block ion permeation in most dilated TMEM16A states.
A) Traces of chloride z-positions in the TMEM16A pore during simulations of the cluster 12 medoid under applied voltages. The grey bar in each plot indicates when a lipid phosphorous atom is within 5 Å of the R515 sidechain. B) Snapshots from simulation of cluster 12 medoid with 300 mV applied membrane potential. Images correspond to timepoints indicated by pink arrows in (A). POPC lipids are shown in yellow and chloride is shown as a green sphere. (*Figure caption continued on the next page.*)

(*Figure caption continued from the previous page.*) C) Snapshots of the cluster 12 pore from simulations with a -300 mV. Images correspond to timepoints indicated by blue arrows in (A). D) Z-positions of chloride during a simulation of cluster 11 with an applied 300 mV potential. E) Snapshots from simulation of cluster 11 medoid with 300 mV applied membrane potential. Images correspond to timepoint indicated by the black arrow in (D).

Discussion

This work is the first to show multiple spontaneous chloride permeation events, at least one in both directions, through the TMEM16A pore without an applied voltage across the membrane. However, several years ago, Jia & Chen [50] used AAMD to sample a singular Cl^- permeation event through a different open conformation of TMEM16A. These previous simulations were initiated from a Ca^{2+}-bound structure (PDB ID 5OYB) that lacks the rotation of TM3 seen in the inhibited state. Nonetheless, like the inhibited state, 5OYB does not have a pore wide enough to accommodate a Cl^-, but when they simulated with PIP_2 bound to the protein about half of their pores became dilated. My simulations clearly show that if your start from the inhibited state and remove 1PBC, each TMEM16A subunit can independently open or constrict the pore and PIP_2 is not necessary to sample the open state. Although the upper half of TM4 appears closer to TM6 in the open states I've identified (7ZK3*[7,11 and 12]) compared to 5OYB* they all have a separated hydrophobic gate which is likely the most critical requirement for getting ions through (**Figure 2.7*A***, **Supplemental Information Table 2.1**). To better compare the behavior of 5OYB* to my simulated open states I simulated 5OYB* with PIP_2 for 1 μs using the Gromacs engine (Jia & Chen used Amber14) and same distance restraints used for the 7ZK3 simulations. During the simulation the pore did not collapse consistent with the original publication but the extracellular ends of TM3 and 4 tilted away from the dimer interface while keeping TM4 straight and the flux of water in the center of pore fluctuates throughout the simulation (**Supplemental Information Figure 2.16*A*)**. Overall, the amount of water moving through the center region of the pore seems less than when the initial 5OYB* conformation is held by 1 kcal/mol·A^2 backbone restraints (**Supplemental Information Figure 2.16*B*)**. I also observed a

lipid head group in the outer pore entrance in the restrained simulation while Jia & Chen only report lipid tails entering parts of the pore (**Supplemental Information Figure 2.16*B, right***).

5OYB* has a larger mean pore radius (~2.25 Å) at the hydrophobic gate than 7ZK3*[11] (~1.8). Both 5OYB* and 7ZK3*[11] also have conductance values within the lower half of experimental estimates: 0.76 pS and ~1.3 pS respectively versus 0.5-8 pS in experiments [83], the larger value for 5OYB* makes sense given its slightly larger pore size. The permeating ion in the 5OYB simulations spends most of its crossing time interacting with K645 which is consistent with the longer dwell times I see in my simulations (**Supplemental Information Figure 2.15**). All together there does not seem to be any functional or mechanical difference between the 7ZK3* and 5OYB* conformations which suggest that the TM3 rotation, does not greatly affect how ions permeate in open states.

In addition to seeing the pore spontaneously open I also saw a similar number of subunits collapse their pore once 1PBC was removed. One of the constricted states in my simulations, cluster 2, is very similar to the Ca^{2+} -bound mTMEM16(a) structure (PDB ID 6B8I [67]) which authors believe represent a rundown state of the protein (**Figure 2.7*B***). 5OYB is similarly thought of as a rundown state and can be made conductive by binding PIP_2 as demonstrated by Jia & Chen. Does this mean that PIP_2 could also recover conductive states from the closed 7ZK3* state or prevent the pore from collapsing? To address this question, I ran one 1 μs simulation of initiated from the inhibitor-bound state with PIP_2 placed at the same location as in the Jia & Chen simulations (**Supplemental Information Figure 2.16*A***). To better compare to the 5OYB* results, I also ran a simulation with identical restraints on parts of the cytosolic domain as in Jia & Chen and a simulation where I removed the third bound Ca^{2+}, which is absent in 5OYB, from the dimer interface (**Supplemental Information Figure 2.16*B, C***). In the simulation the usual distance restrains and PIP_2 both subunits collapsed the pore with the usual bend in TM4 at E555 and dehydration at the center of the pore (**Supplemental Information Figure 2.16*A***). For comparison only 1/6 subunits of the original symmetric model simulations (without PIP_2) collapsed so completely. Half of the remaining subunits (either with different restraints or third Ca^{2+}-removed) also

collapsed their pores (**Supplemental Information Figure 2.16*B, C***). More sampling is needed to draw any strong conclusions about whether PIP_2 alters the open probability in simulations, but these preliminary results indicate that having PIP_2 bound does not guarantee an open channel.

The extent to which PIP_2 influences TMEM16A pore dilation is further complicated by the potential influence of the starting position of cytosolic loops that were missing in the cryo-EM structure and modeled into the protein for simulations if 7ZK3 with 1PBC removed (**Supplemental Information Figure 2.12*B)***. Understanding the reason behind this could be critical to understanding what roles these dynamic loops play in TMEM16A activity and will require more detailed analysis.

Finally, one finding that's difficult to rationalize from my results is the large maximum conductance value for cluster 12 at 300 mV compared to -300 mV. This could be an artifact of the strong membrane potential, which is much higher than physiological values, but this is difficult to determine since experiments only use a smaller range of voltages. Nonetheless, given that this state can also be blocked by lipids I wonder if this is a relevant conductive state at all. Of the clusters I chose to simulate, cluster 12 is the smallest by population and the only one with a disengaged TM3-4 hydrophobic network which. The conformation represented by cluster 11 on the other hand is better sampled by the simulations and has a conductance value in the range, although on the lower end, of experimental estimates. It seems more likely that cluster 11, rather than cluster 12 is the major conductive state of TMEM16A with a rotated TM3.

To conclude, I have used AAMD simulations to sample novel TMEM16A confirmations starting from the 1PBC-inhibted structure after 1PBC was artificially removed from the binding site. The simulations spontaneously split into two major groups: constricted and open. I observed 6 total permeation events in different simulations at zero mV, but only one once the pore sufficiently dilated. Open states derived from the inhibited state tend to keep some part of their TM4 and 6 in close contact by pending or titling TM4 toward TM6 (Figure 2.6). At least two open states far from the starting cryo-EM coordinates, have linear currents and conductance values comparable to experiments. However, the

electronic behavior is complicated by occasional but long-lasting lipid blocking of the ion pore (**Figure 2.6**). Lipids lining the ion permeation path and even temporarily blocking the ion pore is something that has been demonstrated for scramblase TMEM16 members ([44] and detailed further in Chapter 3) so it seems plausible that lipids would also participate in the TMEM16A pore when partially exposed to the membrane.

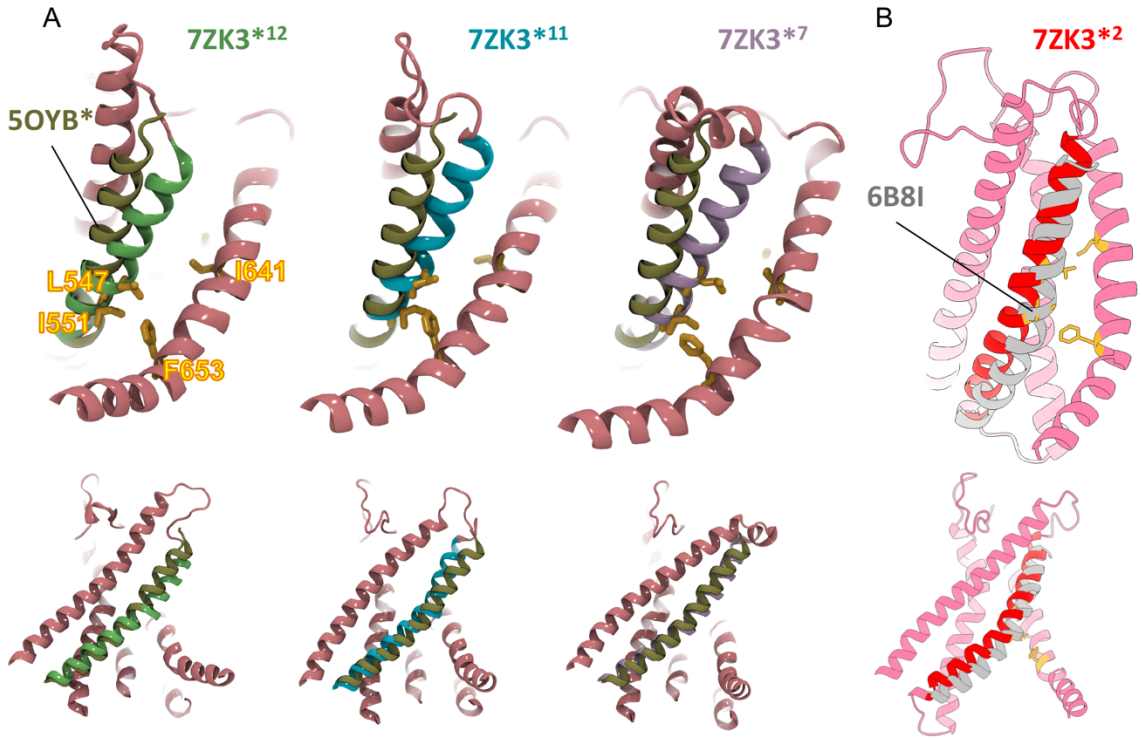

Figure 2.7 Ion conductive states predicted from 1PBC/Ca^{2+}- and Ca^{2+}-bound TMEM16A differ in position of TM4.
A) Cartoon representations of cluster 12 (7ZK3*[12], A), cluster 11 (7ZK3*[11], B) and cluster 7 (7ZK3*[7] ,C) medoids from tICA and MSM analysis on simulations of TMEM16A with 1PBC removed with TM4 colored differently in each. TM4 from predicted open state of TMEM16A (5OYB*) initiated from Ca^{2+}-bound TMEM16A (PDB ID 5OYB) overlayed (olive green). B) Cartoon representations of cluster 2 (7ZK3*[2], A) medoid (pink and red) with TM4 from cryo-EM structure of Ca2+-bound mTMEM16A(ac) (PDB ID 6B8I [67], grey).

Methods

TMEM16A simulation system preparation

All atomistic simulations were initiated from either the Ca^{2+}/1PBC-bound mTMEM16A cryo-EM structure (PDB ID 7ZK3, [51]) or extracted simulation snapshot from simulations initiated from Ca^{2+} -bound mTMEM16A (PBD ID 5OYB) performed by Jia & Chen (see [50] for simulation details). The missing residues in 7ZK3: 260-266, 467-482, 526-527, 669-682 were built using MODELLER (version 10.2 [84]) or extracted from mouse TMEM16A prediction deposited in the AlphaFold Protein Structure Database [77]. MODELLER -predicted loops were refined using the LoopRefine [76] method with 10 iterations per residue. Loops extracted from the AlphaFold prediction where fit to the 7ZK3 after aligning TM7-8 of each structure. These loops were then annealed into the 7ZK3 structure by applying LoopRefine only to 3-4 resides bridging the inserted loops. PROPKA3 was used to check the protonation state of protein residues. E624 and D405 are both weakly protonatable at neutral pH, but likely well solvated and therefore left in their negatively charged states [85]. Each N and C termini of each protein was capped by methylamide and acetyl groups and then embedded in a 155x155 $Å^2$ 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (POPC) bilayer and solvated in 150 mM KCl and the CHARMM TIP3P water model using CHARMM-GUI's Membrane Builder [86]. System charges were neutralized using the same ions. Parameters for 1PBC were generated using the VMD Force Field Toolkit Plugin [87] and Gaussian09 [88].

MD simulations

Simulations were performed with Gromacs (version 2020.6 [89]) and the CHARM36 [90] and CHARMM36m [91] force fields for lipids and protein respectively. During minimization, equilibration and production distance restraints with 418.4 kJ mol^{-1} nm^{-2} force constants were applied between the Cα atoms of residues 465 and 489, 454 and 566, 169 and 278, 126 and 176, 196 and 189, 123 and 282, 185 and 200 to stabilize the cytosolic domain. Simulations were run using a 2 fs time step in an NPT ensemble. Temperature was kept at 303.15 K using the Nosé-Hoover [92] thermostat (τ_T=1 ps). The

pressure of the system was semiisotropically coupled to a 1 bar reference pressure by the Berendsen [93] and Parrinello-Rahman [94] barostat (τ_P=5 ps, compressibility=4.5x10^{-5}) or equilibration and production respectively. All bonds to H were constrained with LINCs [95]. Particle mesh Ewald [96]-calculated electrostatic and Van der Waals (VdW) interactions were cut off at 1.2 nm. A Verlet cut-off scheme was used for non-bonded interactions. VdW interactions were smoothly switched to zero between 1.0 and 1.2 nm. The protein with its bound Ca^{2+} ions, the membrane, and solvent bath were treated as separate groups for the thermostat coupling and center of mass removal. Harmonic restraints of the protein backbone, sidechains, lipids and dihedrals were applied and slowly reduced over 8 equilibration steps totaling ~32 ns. The equilibrated box size was ~143x143x148 Å^3. In simulations with restrained backbones a 418.4 kJ mol^{-1} nm^{-2} harmonic restraint was applied to backbone heavy atoms of TM3 (residues 485-525), TM4 (residues 529-569), TM5 (residues 72-602), TM6 (residues 629-668) and TM7-8 (residues 699-745).

MD simulations with an applied voltage

Voltages was increased in 10 mV jumps every 5 ns starting from the end of normal equilibration until the target voltage was reached. The potential was set using a constant electric field protocol where $E_{applied} = \frac{V}{L_z}$ where V is the voltage and L_z is the system box height (152 Å in these simulations). The applied electric field for a 300-mV potential was 0.0194 $kcal/mol \cdot Å^2 \cdot e$.

tICA and clustering analysis

TM4/5/6 residue pair distances from aggregate atomistic trajectories of TMEM16A were submitted to time-independent component analysis (tICA) [80] and subsequent MiniBatch K-means clustering. 500 clusters were then used to construct a Markov-state model and microstates were grouped into macrostates using the improved Perron-cluster cluster analysis (PCCA+) method [81], [82]. The above analysis was performed using MSMBuilder [97].

Water and ion density calculations and pore pathway generation

Density of water was collected in a 25x35x150 Å^3 box with 0.5 Å uniform length grids centered on L547 in each subunit. At every 2D row in z a geometric mean of non-zero bin coordinates was

calculated weighted the density value in each bin. A 3D spline was then fit the average coordinates. Density of ions was calculated in the same way and its density was projected onto the average water density path by matching each density grid to the nearest path node.

Channel current and conductance measurements

The current for 200 ns blocks of simulation time was calculated using the following equation: $I = \frac{N_{Cl} \cdot e}{t}$ where N_{Cl} is the number of complete permeations events in the time block. The conductance was calculated by dividing the current, I, by the membrane potential. Average and standard deviations of currents from each widow was reported as the mean and error in Figure 2.4.

Contact, dwell time and protein feature analysis

At every simulation time step the positions of all water and ions are recorded if they fall within a 12 Å radius of the average water density pathway and binned into 3.5 high cylinders along the same pathway. Contact was also recorded for any protein residue that came within 3.5 Å of an ion with the pore cutoff distance. Interaction lifetimes were recorded for any continuous contact with a unique ion. All of this analysis and measurements for TM kink angles and residue distances were done using custom python scripts with on MDAnalysis [98] and SciPy [99] methods. Pore radii were calculated using the HOLE2 [100] implementation in MDAnalysis.

3. GROOVES, INTERFACES, AND OTHER SURPRISING WAYS LIPIDS TRANSVERSE THE BILAYER VIA TMEM16

Introduction

It was first theorized [21][101] and later predicted by molecular dynamics (MD) simulations [37][38][42][14][54][57] that lipids can transverse the membrane bilayer by moving their headgroups along the water-filled hydrophilic groove (between TM4 and 6) while their tails project into the greasy center of the bilayer. This mechanism for scrambling, first proposed by Menon & Pomorski [102], is often referred to as the credit card model. All-atom MD (AAMD) simulations of open *Nectria haematococca* TMEM16 (nhTMEM16) have shown that lipids near the pore frequently interact with charged residues at the groove entrances, two of which are in the scrambling domain which confers scramblase activities to ion channel TMEM16A chimeras. Frequent headgroup interactions with residues lining the groove were also noted in atomistic simulations of open TMEM16K including two basic residues in the scrambling domain. Lipids experience a relatively low energy barrier for scrambling in open nhTMEM16 (<1 kcal/mol compared to 20-50 kcal/mol through the raw bilayer) [38][102]. Simulations also indicate that zwitterionic lipid headgroups stack in the open groove along their dipoles which may help stabilize them as they cross [38][44]. Finally, simulations also show that lipids can directly gate nhTMEM16 groove opening and closing through interactions with their headgroups or tails [41][40]. Lipids are also directly involved in how TMEM16 scramblase conduct ions. As first speculated in [13], AAMD simulations have shown that ions permeate through the lipid headgroup-lined hydrophilic groove [39][44][60][43][41]. So far, this mechanism has only been demonstrated for TMEM16K and nhTMEM16 in atomistic simulations. AAMD simulations have even shown that the certain types of lipids headgroups can transiently block ion permeation, toggle ion selectivity and modulate free energy barriers for cation permeation in nhTMEM16 [43] [44]. Free energy barriers for K^+ permeation can vary by as much as 5 kcal/mol depending on which lipids occupy the pore [43]. Unlike other scrambling members, nhTMEM16 can adopt intermediate conformations [65] that can conduct ions but not scramble lipids as

efficiently as the open state [65]. However, AAMD simulations have shown that lipid headgroups can still partially line the intermediate groove and form part of the ion pore [44]. Simulations of TMEM16A, show that groove opening, although not as wide as open TMEM16 scramblases, is also required for Cl^- permeation but lipid headgroups do not form part of the pore like for the TMEM16 scramblases [50] indicating that TMEM16 ion channels and TMEM16 scramblases use different mechanisms to conduct ions. However, all of these simulation observations have been on a limited number of spontaneous events from different groups (in aggregate we estimate up to 14 scrambling events and only two ion permeations in the absence of an applied voltage) [14] [38] [42] [39] [60] [54] [57]. Many more scrambling events (~800) have been seen in coarse-grained MD (CGMD) simulations for nhTMEM16 [37], TMEM16K [14], mutant TMEM16F (F518H) and even TMEM16A [103], however a detailed analysis of how these scrambling events occurred is missing for the latter two.

One of the biggest uncertainties in the TMEM16 field now is how TMEM16F scrambles lipids. So far only closed conformations (grooves too constricted to fit lipid headgroups) have been solved experimentally [71][72]. This has caused the field wonder if an open groove is even necessary for scrambling. This idea is further supported by liposome-based scrambling assays showing that fungal and mammalian TMEM16s scramble lipids in the absence of Ca^{2+} when their grooves are presumably closed [13], [45][21][14][15][42][104] and that afTMEM16 can scramble PEGylated lipids which are too large to fit even through the open groove [45]. This last result motivated Malvezzi *et al.* [45] to propose that lipids can be scrambled outside of the groove (here on referred to as “out-of-the-groove” scrambling). This mechanism leans on observations from cryo-EM nanodiscs [65][62] [71][64], MD simulations [38][40][14] and continuum models [38] that show that all TMEM16 members locally distort the membrane when the groove is open and closed, although the membrane defects are more sever around open fungal TMEM16s [62] [65] and some TMEM16F [71] structures. The membrane near the fungal TMEM16 and TMEM16F groove is also thinner (between 50 and 60%) than the normal bulk bilayer which in addition to the membrane packing defects could lower the energetic barrier to lipid crossing

[38], [62][71]. In this mechanism the groove still provides a conduit for lipids to scramble in both closed and open states, but lipids do not *need* to insert their headgroups, per the credit card model, due to the presence of the membrane distortion. While no AAMD simulations have reported scrambling by closed TMEM16F or by any closed TMEM16, a recently published CGMD simulation of the closed F518H TMEM16F mutant did have some scrambling events, but its mechanism was not specified [103]. Since a comprehensive analysis across all family members has not been carried out it is difficult to determine how membrane thinning is related to scrambling or if scrambling mechanisms are specific to certain family members, conformational states of the protein or both.

To address these outstanding questions, we employed CGMD simulation to systematically quantify scrambling in 27 experimental and computationally predicted TMEM16 proteins taken from each family member that has been structurally characterized: nhTMEM16, afTMEM16, TMEM16K, TMEM16F, and TMEM16A (**Supplementary Information Table 3.1, Supplementary Information Figure 3.1**). CGMD, which was the first computational method to identify nhTMEM16 as a scramblase, enables us to reach much longer timescales, while retaining enough chemical detail to faithfully reproduce experimentally verified protein-lipid interactions [105]. This allowed us to quantitatively compare the scrambling statistics and mechanisms of different WT and mutant TMEM16s in both open (including predicted) and closed states and solved under different conditions (e.g. salt concentrations, lipid and detergent environments, in the presence of modulators or activators like PIP_2 and Ca^{2+}). Our CG MD simulations successfully reproduce experimentally determined membrane deformations seen in nanodiscs across both fungal and mammalian TMEM16s. Among individual scramblase family members only structures that are both open and Ca^{2+}-bound have grooves fully lined by lipids, but for TMEM16F this was only seen for a mutant and predicted open state. All of these structures promote scrambling in the groove with lipids experiencing a less than 1 kT free energy barrier as they move between leaflets. One simulation of TMEM16A, which is not a scramblase, initiated from a predicted ion conductive state also has lipids across its entire groove and had two scrambling events, but its maximum lipid free energy

barrier was 2.5-5.5-fold larger than that of the scramblases. Our analysis of the membrane deformation and groove conformation shows that most scrambling in the groove occurs when the membrane is thinned to a least 8 Å and the groove is open. Our simulations also reveal alternative scrambling pathways including a single out-of-the-groove event for Ca^{2+}-bound TMEM16F, but more frequent alternative pathways for mammalian family members occur at the dimer interface between TM3 and TM10 from opposite subunits.

Results

CG simulations of TMEM16 scramblases only have lipids in the open groove

We simulated coarse-grained (CG) Ca^{2+}-bound structures of TMEM16 proteins in a 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) bilayer for 10 μs each using the Martini3 [106] forcefield and restricted our analysis to the last 9 μs to account for system equilibration. First, we determined how well the simulations reproduced the experimentally determined membrane shapes by comparing the annulus of lipids that surrounds each protein to the lipid density of TMEM16s structures solved in nanodiscs (**Supplementary Information Figure 3.2**). Overall, the shape of the membrane near the protein are qualitatively very similar to the experimental densities and AAMD simulations and continuum model predictions [38], [62], [65]. For example, the CG simulations also capture the sinusoidal curve around both fungal scramblases in apo and Ca^{2+} -bound states (**Supplementary Information Figure 3.2**). The weak density above TM7-8 from the simulation of the Ca^{2+} -bound mutant TMEM16F also corresponds well to the lack of cryo-EM nanodisc lipid density in the same location (**Supplementary Information Figure 3.2**). This CG deformation pattern is also consistent with our atomistic simulations [38] of Ca^{2+}-bound nhTMEM16 using the CHARMM36 forcefield (**Figure 3.1A-*B***).

Single snapshots and density isosurfaces from CGMD simulation of known scramblases nhTMEM16 (PDB ID 4WIS), afTMEM16 (PDB ID 7RXG), TMEM16K (PDB ID 5OC9) and constitutively active mutant TMEM16F F518H (PDB ID 8B8J) show that lipid headgroups occupy the full length of the groove (**Figure 3.1*B***). In each of these simulations the groove is also filled with water

(**Supplementary Information Figure 3.3**, **3.16**). Arndt *et al.* [64] showed that a F518H mutation in TM4 turns TMEM16F into a constitutively active scramblase that is structurally characterized by a kink in TM3 and a rearranged TM4-TM6 interface. To date, the scrambling mechanism of this mutant – and whether it resembles the activity of the WT protein – is poorly understood. Our simulations suggest that lipids insert their headgroups into the solvated groove in a nearly identical manner as the fungal scramblases and TMEM16K, consistent with the "credit card model" for scrambling (**Supplementary Information Figure 3.3*A-B***). During the simulation the F518H TMEM16F mutant groove also dilated slightly which will be detailed later. Unlike the Ca^{2+}-bound scramblase structures, a simulation of one of our predicted conductive states of Ca^{2+}-bound TMEM16A (7ZK3*[6] , see Methods and **Supplementary Information Figure 3.1**) samples some lipid headgroups entering the extracellular vestibule formed by TM3-6, but the center of the potential pathway appears inaccessible to lipids as indicated by the lack of density near the center of the membrane (**Figure 3.1*B***). This supports experimental evidence that TMEM16A lacks scramblase activity [107][56] although later in this paper we will show that a different TMEM16A conformation has a fully lipid-line groove similar to the scramblases. Density isosurfaces from simulations of scramblase apo or closed states (**Supplementary Information Figure 3.4**, **3.16**) also show a lack of lipid headgroup density near the center of the bilayer.

In each simulation of the Ca^{2+}-bound open scramblases, there was a clear upward deflection of the membrane as it approaches TM3, 4 and the extracellular TM5-6 helix from the left and downward deflection as it approaches TM6 and 8 from the right (**Figure 3.1*B***). However, we noted a difference in the strength of lipid density near the extracellular groove entrances where the membrane approaches TM6. In nhTMEM16 and TMEM16K simulations there appears to be a strong presence of lipid headgroups across TM6, TM2 and above the TM7-8, but this density appeared weaker at the equivalent location for the two other scramblases afTMEM16 and TMEM16F. The density from simulations of apo structures appeared more consistent across the family at the same location (**Supplementary Information Figure 3.4**). Lipids have been resolved at the same site in cryo-EM structurers of fungal scramblases in

nanodiscs [61][66] and residues in this site on nhTMEM16 and TMEM16F groove also seem to play a role in scrambling but the mechanism by which they do so is unclear [62], [66] [72][71].

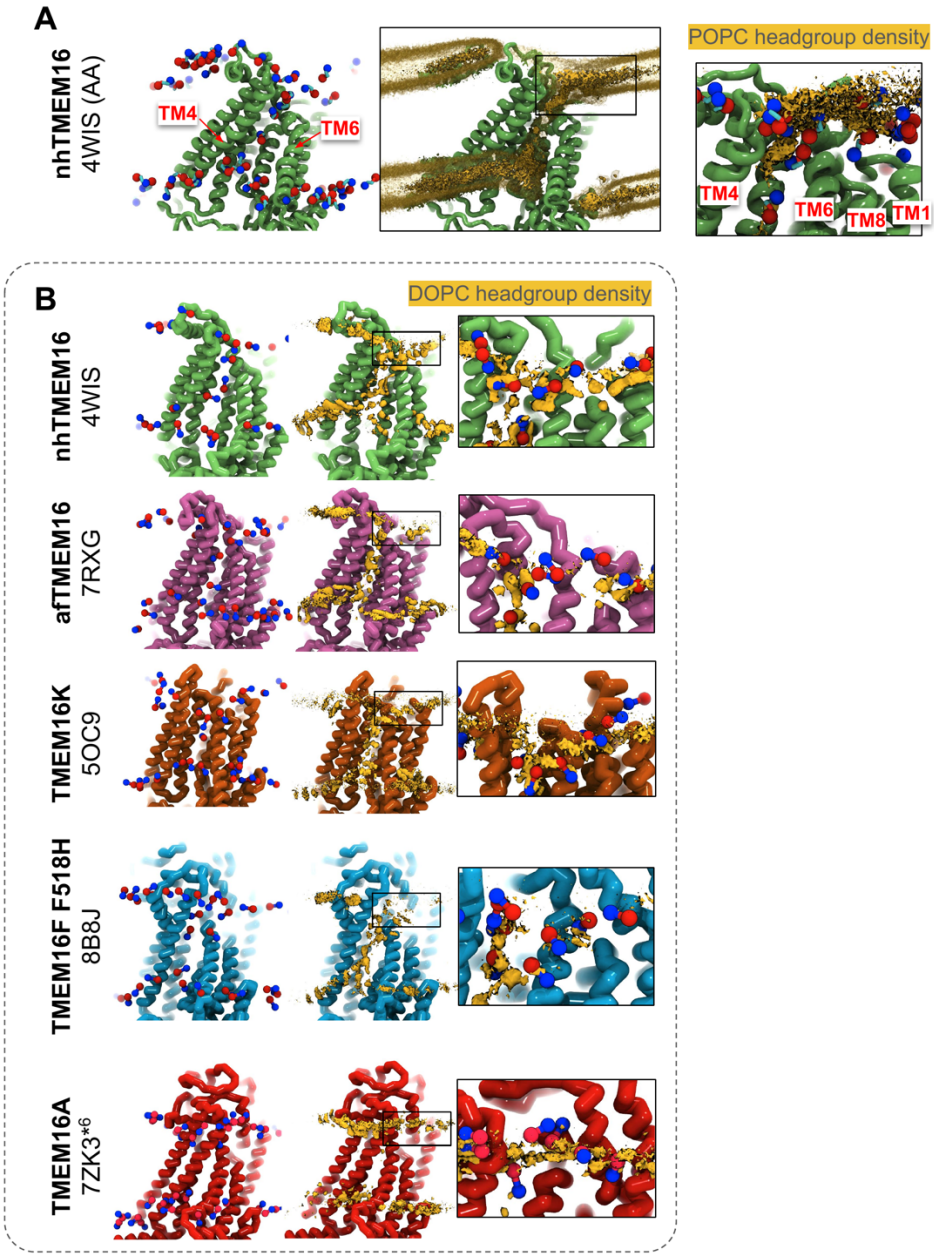

Figure 3.1 CG simulations of multiple TMEM16 structures captures lipid density in the TM4/6 pathway of scrambling competent members.
A) a snapshot and POPC headgroup density (right) from atomistic simulations of Ca^{2+}-bound nhTMEM16 (PDB 4WIS) previously published in Bethel & Grabe 2016 [38]. Only the PC lipid headgroup is shown for clarity. Density is averaged from both subunits, except TMEM16K 50C9 which is asymmetric, across 8 independent simulations totaling ~2 μs. (*Figure caption continued on the next page.*)

(*Figure caption continued from the previous page.*) B) Snapshots from coarse-grained simulations of nhTMEM16 (green), afTMEM16 (violet), TMEM16K (orange), TMEM16F (blue), TMEM16K (orange), and TMEM16A (red). Each density is averaged over both chains except TMEM16K and TMEM16A where only a single chain is used due to the structure's asymmetry. Only the PC lipid headgroup is shown for clarity. Density was calculated using the last 9 μs of simulation time.

Simulations recapitulate scrambling competence of open/closed structures

To quantify the scrambling competence of all simulated TMEM16 structures, we analyzed individual lipid angles with respect to the membrane normal and counted the number of events in which these angles transition from ~30° (typical for the lower leaflet) to ~150° (typical for the upper leaflet), or *vice versa* (see Li et al ([108]), **Supplementary Information Figure 3.5** and Methods for details). The scrambling rates calculated from our CGMD trajectories are in excellent agreement with the open/closed configurations described for the experimental structures (**Figure 3.2*A***). The most scrambling competent structure in our assay was the Ca^{2+}-bound X-ray structure for fungal nhTMEM16 (PDB ID 4WIS [21]), with 24.44 ± 5.23 events/μs. In line with experimental findings [65], the open Ca^{2+}-free structure (PDB ID 6QM6, solved by cryo-EM in detergent) that is structurally very similar to 4WIS also scrambled lipids in our simulations (15.67 ± 3.94 events/μs). In contrast, we observed no scrambling events for the "intermediate" (PDB ID 6QMA) and "closed" (PDB ID 6QM4, PDB ID 6QMB) structures from the same study [65]. We observed a similar trend for the fungal afTMEM16, where our simulations correctly identified the open Ca^{2+}-bound cryo-EM structure (PDB ID 7RXG) as a scrambling competent conformation while the Ca^{2+}-free closed-groove structure (PDBID 7RXB) from the same study [61] did not scramble.

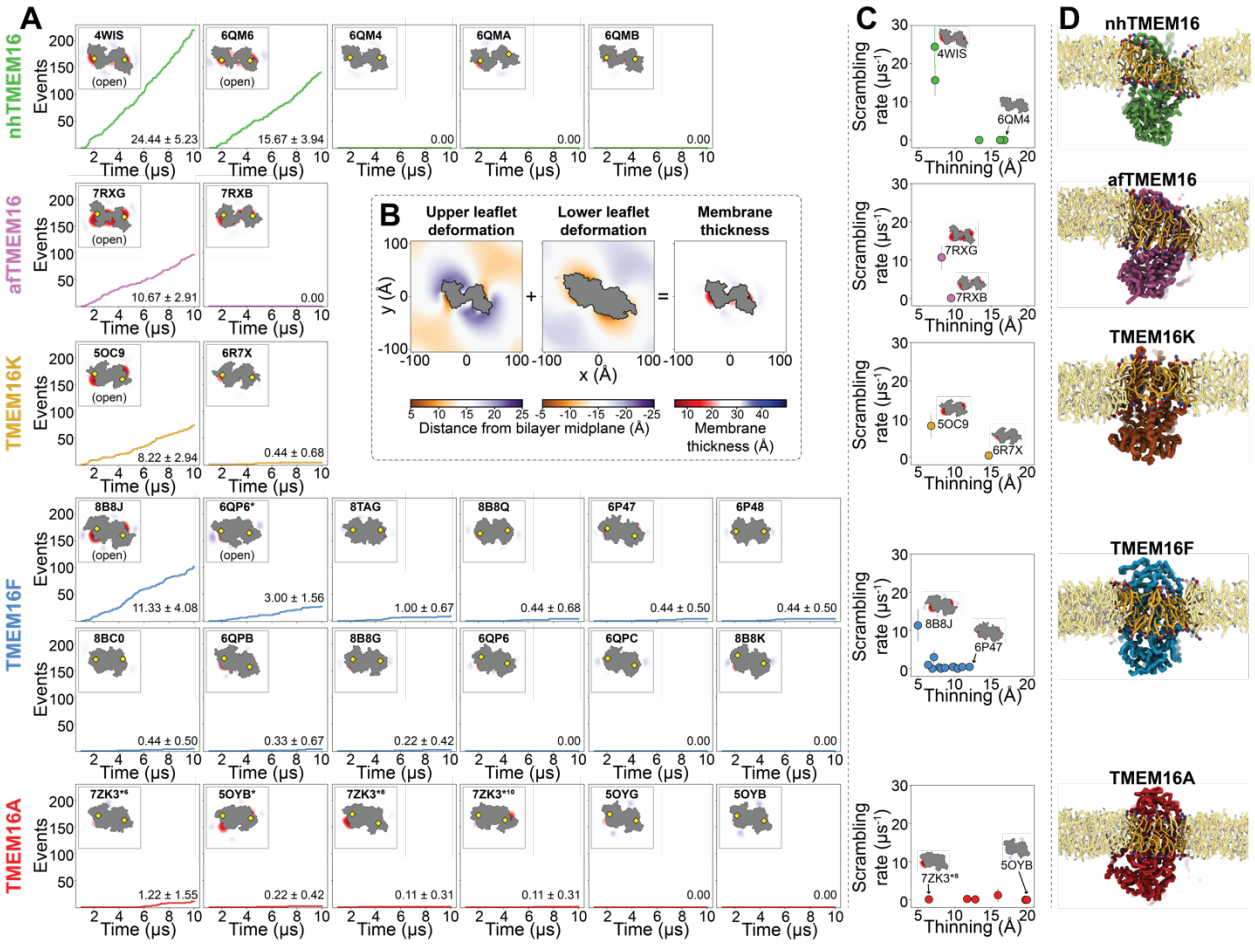

Figure 3.2 Simulated lipid scrambling differentiates closed/open conformations and correlates with membrane thinning.
A) Accumulated scrambling events over 10 μs of CGMD simulation of experimental and simulated (sim) structures of nhTMEM16 (green), afTMEM16 (violet), TMEM16K (gold), TMEM16F (blue), TMEM16K (orange), and TMEM16A (red). Average and standard deviation of scrambling rates are calculated by block-averaging over the last 9 μs of simulation time (1 μs blocks). Insets show 120x120 Å membrane thickness plots (see colorbar in *B*). The protein cut-out (gray) is a top-view representation of the portion of the protein above the membrane midplane (z>0) taken from the starting frame (t=0). Yellow diamonds indicate the position of groove (backbone bead of central residue in TM6; Y439, Y432, T435, I611, I640 for nhTMEM16, afTMEM16, TMEM16K, TMEM16F, and TMEM16A, respectively. B) The local membrane thickness (parallel to the membrane normal) was calculated as the sum of the membrane deformations of the upper and lower membrane leaflets. Only 4WIS is shown here as an example. All membrane deformation plots are shown in SI Fig. X. C) Correlation between the minimal membrane thickness (the thinnest points in the insets in Fig. 2A) and scrambling rate. D) Simulation snapshots (side-view) of the most scrambling-competent protein structures for each family member.

For TMEM16K, our simulations showed that the Ca^{2+}-bound X-ray structure (PDBID 5OC9) is scrambling competent, in line with the experiments and MD simulations described in the original work

[14]. Interestingly, we found a significant asymmetry in the number of scrambling events between the two different monomers, with >80% of the 8.22 ± 2.94 event/μs events happening on "chain B" (**Supplementary Information Figure 3.7*A***). Although both monomers are Ca^{2+}-bound, the chain B has a slightly wider ER lumen entrance to the groove (**Supplementary Information Figure 3.7*B***). As a closed TMEM16K conformation, we simulated the cryo-EM structure PDB ID 6R7X from the same paper [14], which indeed showed very little scrambling activity (0.44 ± 0.68 event/μs) in our simulations. Although TMEM16F is known to act as a lipid scramblase in the plasma membrane [12], none of the many WT protein structures that are available to date are in an open state, including constitutively active mutants like TMEM16F F518H. In our simulations, F518H TMEM16F (8B8J) was indeed the only experimental structure that showed scrambling activity. We also performed a CGMD simulation starting from a partially open configuration of WT TMEM16F that was obtained by atomistic MD simulation of a closed-state structure (6QP6) with enhanced sampling (cluster 10 in [57], 6QP6* in Figure 3.2*A*). This structure features one monomer in the open state and one monomer in the closed state. In accordance with this, we observed moderate lipid scrambling activity (3.00 ± 1.56 events/μs) in the open groove and none in the closed (**Supplementary Information Figure 3.8*B-C***).

Finally, we simulated several structures of mouse TMEM16A, which functions as an ion channel but lacks lipid scrambling activity. In agreement with the experiments [63], both the Ca^{2+}-bound (5OYB) and the Ca^{2+}-free (5OYG) experimental structures were indeed inactive in our simulated lipid scrambling assay. Ion conductive structures that were obtained by atomistic MD by our own group (see Methods for details)) had very little scramblase activity. A dilated TMEM16A state predicted by Jia & Chen (5OYB*) [50] however scrambled a single lipid trough each groove in a manner nearly identical to the scramblases (**Supplementary Information Figure 3.9*A***).

Our CGMD simulations correctly differentiate between open and closed conformations across the five TMEM16 family members we studied here. This notion is in line with the work of Li et al ([108]), who also showed good agreement between *in vitro* and *in silico* lipid scrambling using the same Martini 3

force field on a diverse set of protein structures. Because the simulation conditions and system setups were identical in all our simulations, the recent study allows for a unique direct comparison of scrambling rates between different TMEM16 structures.

Scrambling rate correlates with membrane thinning and groove dilation

As described previously – both in cryo-EM and simulation studies ([38], [64], [71][14], [17], [61])– membrane thinning is suggested to contribute to the scramblase activity of TMEM16s. To investigate such membrane deformations in our simulations, we calculated the ensemble-averaged positions of the glycerol groups for all lipids in our systems and plotted the position along the z dimension (the membrane normal) across the xy-plane (see methods for details). For most of the TMEM16 structures, we indeed observed strong deformations close to the protein surface, with opposite signs for the upper and lower leaflets (left two panels in **Figure 3.2*B***, **Supplementary Information Figure 3.10-14** for all data). Summing the two deformation maps yields an xy-plot of the local membrane thickness (right panel in **Figure 3.2*B*** and insets in **Figure 3.2*A*** and **Figure 3.2*C***, all data in **Supplementary Information Figure 3.10-14**). This visualization indicates that active scramblases indeed tend to feature more extreme membrane thinning, as indicated by larger and deeper red areas in the insets in **Figure 3.2*A*** (e.g., 4WIS, 7RXG, 5OC9, and 8B8J). Interestingly, the thinning tends to be the strongest near the hydrophilic groove of the protein that facilitates lipid scrambling (yellow diamonds in Figure 3.2*A*), in line with previous studies [38] [61].

We quantified this correlation by plotting the scrambling rates against the minimal membrane thicknesses (i.e., the thinnest points in the respective membrane thickness maps, Figure 3.2*C*). For all four scrambling-competent TMEM16 families, the active scramblases indeed showed more potent membrane thinning than the inactive structures. This trend is preserved when excluding the scrambling lipids from the analysis for the fungal scramblases and TMEM16K (**Supplementary Information Figure 3.15**). As illustrated by the simulation snapshots in **Figure 3.2*D***, we noticed that the Ca^{2+}-bound fungal nhTMEM16 and afTMEM16 structures deform the membrane in a more drastic long-range manner

compared to the active mammalian TMEM16K and TMEM16F structures which induce a sharp local thinning but leave the rest of the membrane relatively flat.

Although membrane thinning is well established for TMEM16, the question of whether it is *sufficient* for scrambling is still an open one [109]. Our results show that most structures that induced thinning below ~8 Å were indeed scrambling competent. However, we did observe similar strong membrane thinning for several closed TMEM16F structures, as well as one of the TMEM16A structures (7ZK3*[8]), without this resulting in lipid scrambling (**Figure 3.2*C***). This finding indicates that those structures may – thinning-wise – be set-up to scramble lipids but are restricted to do so by another determining factor that is likely related to the conformation of the protein itself or its interactions with the direct membrane environment.

Thus far a general feature of TMEM16 structures is a dilation of the upper (extracellular) hydrophilic groove associated with Ca^{2+} binding to the orthosteric site [109]. The exceptions to this are TMEM16A, which is not scrambling-competent, and TMEM16F for which open states have only been observed in MD simulations [57]. These findings strongly indicate that that groove-opening is allosterically coupled to Ca^{2+} binding but do not conclusively determine if an open groove is necessary for scrambling. To quantify how groove openness relates to scrambling we measured the minimum distance between residues on TM4 and TM6 at the most constricted location in the pathway (see water permeation analysis in methods for details). We observed no scrambling for the fungal TMEM16 and only several events for TMEM16K and TMEM16F conformations with mean distances less the 4.5 Å at the groove's constriction point (**Figure 3.3*A***). Given that the bead diameter of the CG DOPC phosphate group is 4.7 Å, the groove needing to be wide enough to allow the headgroup to transverse the path is consistent with the credit-card model.

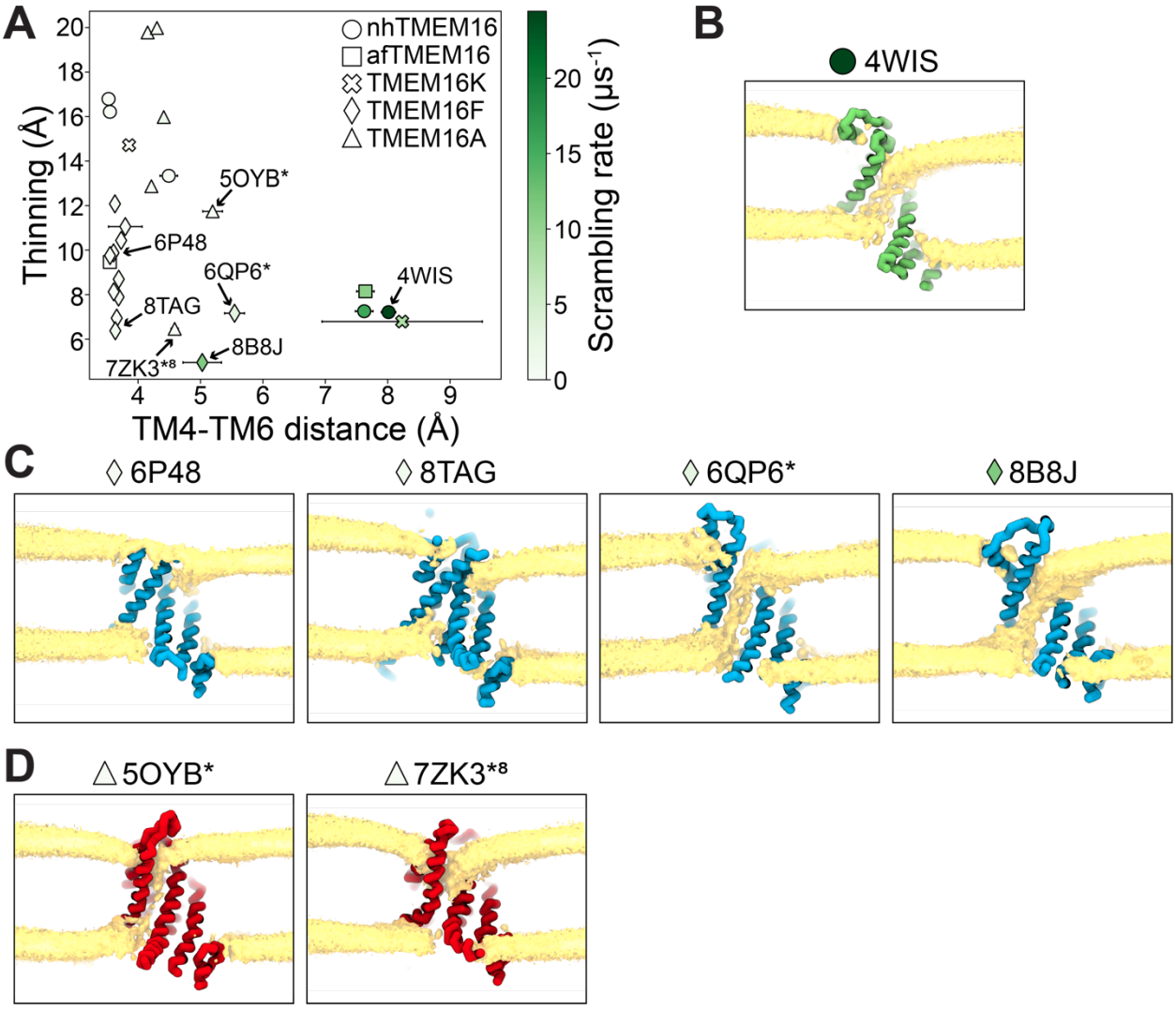

Figure 3.3 Simulated lipid scrambling correlates with groove openness and membrane thinning. (A) The minimum membrane thickness (thinning) plotted against the width of the groove measured by minimum distance between two residues on TM4 and TM6. For asymmetric structures mean distance is only calculated from the subunit with the most scrambling events. Average and standard deviation of TM4/6 distances are calculated by block-averaging over the last 9 μs of simulation time (1 μs blocks). Density isosurfaces for DOPC headgroup beads from (B) nhTMEM16, (C) TMEM16K and (D) TMEM16A simulations. Densities were calculated over the last 9 μs of simulation time. Densities are averaged across both subunits except for asymmetric structures 7ZK3*[8], 6QP6* and 8TAG.

Surprisingly, the TMEM16F F518H mutant (PDBID 8B8J [64]) had scrambling rates comparable to open afTMEM16 and TMEM16K but had a ~0.5 fold less open groove (**Fig. 3*A***). However, this mutant TMEM16F had a nearly 1.5-fold thinner membrane at the groove, which we have shown also positively correlates with scrambling, excluding TMEM16A (**Figure 3.2*C***, **Figure 3.3*A***). The membrane

deformation around the TMEM16F F518H generates the thinnest membrane we observed, even when scrambling lipids are not included in the analysis (**Supplementary Information Figure 3.15**). The membrane profile at the groove entrances look very similar between open nhTMEM16, the simulated open and mutant TMEM16F (6QP6* and 8B8J) in that the upper leaflet is lowered on the TM6 side and lifted around TM4 (**Figure 3.3*B-C***). This suggest that how lipids approach the groove does not seem to be related to their scrambling competence. The membrane profiles near the groove for non-scrambling competent WT TMEM16F structures (6P48 and 8TAG) are also similar to 6QP6* and 8B8J, but they lack lipid density along the full length of the groove, and they do not scramble at the groove (**Figure 3.3*C***). Even though 8TAG can thin the membrane as much as open nhTMEM16 it appears unable to pass lipids along the groove (**Figure 3.3*A***).

One simulated state of TMEM16A (5OYB*) also had a mean distance wider than 5 Å (**Figure 3.3*A***). It has a continuous density of lipids headgroups from the along its open groove (**Figure 3.3*D***) indicating lipids can regularly line the groove even for ion-channel TMEM16s when the groove is sufficiently dilated. We even observed two in-the-groove scrambling events during this simulation (**Supplementary Information Figure 3.9*A***). Its low scrambling numbers compared to mutant and simulated open TMEM16F, despite having similar TM4-6 distances, indicate that additional membrane thinning could increase its ability to scramble. A second predicted TMEM16A conductive state (7ZK3*[8]) on the other hand thins the membrane as much as 6QP6* but does not have a fully lipid-lined groove (**Figure 3.3*D***) and the lipid pathway appears obstructed where TM4 and 6 are closest and we do not observe scrambling in the 7ZK3*[8] groove.

Even when the groove is inaccessible to lipids in closed and intermediates states of TMEM16s we observed water throughout their grooves/pores, however this potential conduit appears shielded from the hydrophilic core (**Supplementary Information Figure 3.3**). As the groove opens the water becomes exposed to the membrane core and we see lipid headgroups insert themselves in the water-filled groove (**Supplementary Information Figure 3.3**), which is consistent with fully atomistic simulations

[37][38][42][39][14][54][41], [43], [57],[44][60]. To quantify how hydration of the groove or pore relates to scrambling we measured the number of water permeation events along the pathway of maximum water density in our CGMD simulation trajectories (**Supplementary Information Figure 3.16**). As expected, the number of water permeation events for most closed structures was low: < 200 events/μs (**Supplementary Information Figure 3.16,** for all data see **Supplementary Information Figure 3.17-22** and **Supplementary Information Data Table 3.1**). In contrast, for structures with more dilated TM4-TM6 grooves (most of them Ca^{2+}-bound) we typically observed 350-500 events/μs (**Supplementary Information Figure 3.16**).

Accordingly, we also recorded spontaneous transport of Na^+ and Cl^- ions (**Supplementary Information Figure 3.17-20**, **22**), in line with the known ion-conductance capacity of these proteins [10][11][12][13][14][15][16][17][43], [44]. Open nhTMEM16 structures (4WIS and 6QM6) sampled multiple ion permeation events and we measured cation to anion event ratios $N_{Na}/N_{Cl} = 5.1$ and $N_{Na}/N_{Cl} = 3.2$ for each simulation respectively. The open afTMEM16 simulation sampled 3x less events than nhTMEM16 but also had poor cation selectivity ($N_{Na}/N_{Cl} = 6$). Both fungal TMEM16s have weak cation (afTMEM16 experimental permittivity ratio $P_K/P_{Cl} = 0.1$, nhTMEM16 simulated $P_{Na}/P_{Cl} = 8.7$) or non-selectivity in experiments and atomistic simulations [15][13][44]. The TMEM16F F518H mutant, which had the most ion permeation events across the family, simulated open TMEM16F (6QP6*) and open TMEM16K simulations have permeation ratios between 0.33 and 2 which is also consistent with their experimentally measured non- or weak cation selectivity (WT TMEM16F $P_{Na}/P_{Cl} = \sim 2.3\text{-}4.8$) [110][14]. Finally, the simulated conductive TMEM16A (7ZK3*[8]) only had 4 Cl^- events which is consistent with its experimentally measured anion, although weak, selectivity ($P_{Na}/P_{Cl} = 0.1$) [47]. Exact permittivity ratios vary depending on ion concentrations and lipid environments but in general TMEM16s are weakly selective or non-selective for ions and that is reflected in our simulations.

As described in the "credit card" scrambling model where a lipid headgroup requires an open groove to translocate from one leaflet to the other, we observed clear correlations between groove

openness and the number of scrambling events in our simulations (**Figure 3.3*A***, **Supplementary Information Figure 3.16**). Taken together with the membrane thinning correlations described earlier (**Figure 3.3*C***), these findings suggest that TMEM16s require both thinning *and* groove opening for scrambling at rates high enough to be detected within our simulation setup. The width of the groove and scrambling rates are also positively associated with the number of water and ion permeation events.

Scrambling also occurs "out-of-the-groove"

Most of the focus on TMEM16 scramblase mechanism has been on the hydrophilic groove formed by TM helices 3-8 which undergoes an obvious conformation change triggered by Ca^{2+} binding to the orthosteric site. In our simulations we see 92% of our scrambling events occur along TM4 and 6 with headgroups embedded in the hydrated groove, in line with the credit card model, which we refer to as "in-the-groove" scrambling (**Table 3.1**). Scrambling events do not appear to enter and leave the groove at specific locations and do not require prolonged contact with the protein outside of the groove prior to scrambling (**Supplementary Information Figure 3.6-9**) with roughly only 3-10% of events passing through high density areas on lower TM4 and upper TM6-8 (**Figure 3.1*B***). Indeed, less than half of the lipid headgroup contact frequency comes from scrambling lipids for residues that we previously identified as being functionally important for organizing lipids in the groove (**Supplementary Information Figure 3.25**). Our calculated free energy profiles of lipids in open grooves show a very low (< 1 kT) energic barrier for fungal TMEM16s, TMEM16K and TMEM16F (**Supplementary Information Figure 3.23**) and kinetics of lipid scrambling appears similar between the homologs with mean diffusion coefficients ~4 $Å^2/ns$ (**Supplementary Information Figure 3.24**). The lipid headgroup interactions with groove-lining residues appear transient with mean dwell times for scrambling lipids less than 1% of the total simulation time (**Figure 3.4, Supplementary Information Figure 3.26-27**). Dwell times are evenly distributed across residues on TM3-7 for the fungal scramblases and TMEM16F, however lipids spend nearly double the amount of time near residues at TMEM16K constriction point on average

(**Supplementary Information Figure 3.25**). Overall, the low interaction times along the groove are consistent with the nearly barrierless free energy profiles (**Supplementary Information Figure 3.23**).

Table 3.1 Number of scrambling events in and out of the canonical groove pathway. Scrambling events where the lipid headgroup transition between leaflets within 4.7 Å of the DOPC maximum density pathway. All other events were considered "out-of-the-groove". For the full list of simulations and scrambling rates see Supplementary Information Data Table 3.1.

homolog	PDB code	# of in-the-groove events	# of out-of-the-groove events	total	average events /μs
nhTMEM16	4WIS	219	1	220	24.44±5.23
nhTMEM16	6QM6	141	0	141	15.67±3.94
afTMEM16	7RXG	96	0	96	10.67±2.91
TMEM16K	5OC9	66	8	74	8.22±2.94
TMEM16K	6R7X	0	4	4	0.44±0.68
TMEM16F F518H	8B8J	98	4	102	11.33±4.08
TMEM16F	6QP6*	24	3	27	3.00±1.56
TMEM16F T137Y	8TAG	0	9	9	1.00±0.67
TMEM16F	6P47	1	3	4	0.44±0.50
TMEM16F	6P48	0	4	4	0.44±0.50
TMEM16F F518/Q623A	8BC0	0	4	4	0.44±0.50
TMEM16F F518H	8B8Q	0	4	4	0.44±0.68
TMEM16F F518H	8B8G	2	0	2	0.22±0.42
TMEM16F	6QPB	0	3	3	0.33±0.67
TMEM16A	7ZK3*[6]	0	11	11	1.22±1.55
TMEM16A	5OYB*	2	0	2	0.22±0.42
TMEM16A	7ZK3*[10]	0	1	1	0.11±0.31
TMEM16A	7ZK3*[8]	0	1	1	0.11±0.31

Surprisingly, ~8% of the remaining scrambling events occurred at the dimer interface with lipids inserting their headgroups into the cavity outlined by TM10 and TM3 (**Figure 3.4, Supplementary Information Figure 3.6-9**). We only observed scrambling at this location in simulations of the mammalian homologs. In atomistic simulations of a closed Ca^{2+}-bound TMEM16F (PDB ID 6QP6) we

observed a similar flipping event for a POPC lipid into the dimer interface (**Supplementary Information Figure 3.28)**. Although the dimer interface is largely hydrophobic there are a few polar and charged residues in the cavity near the membrane core and water is present in the lower half of the cavity (**Supplementary Information Figure 3.29**). In fact, the headgroup of the lipid in our atomistic simulation of TMEM16F interacts with a glutamate (E843) and lysine (K850) on TM10 near the membrane midplane (**Supplementary Information Figure 3.28**). Lipids that scramble at the dimer interface tend to interact for up to 10-fold longer periods of time on average than those in the canonical groove (**Figure 3.4*B***). The most prolonged interactions are enriched at sites with aromatic residues which lipids tails appear to intercalate (**Supplementary Information Figure 3.30**).

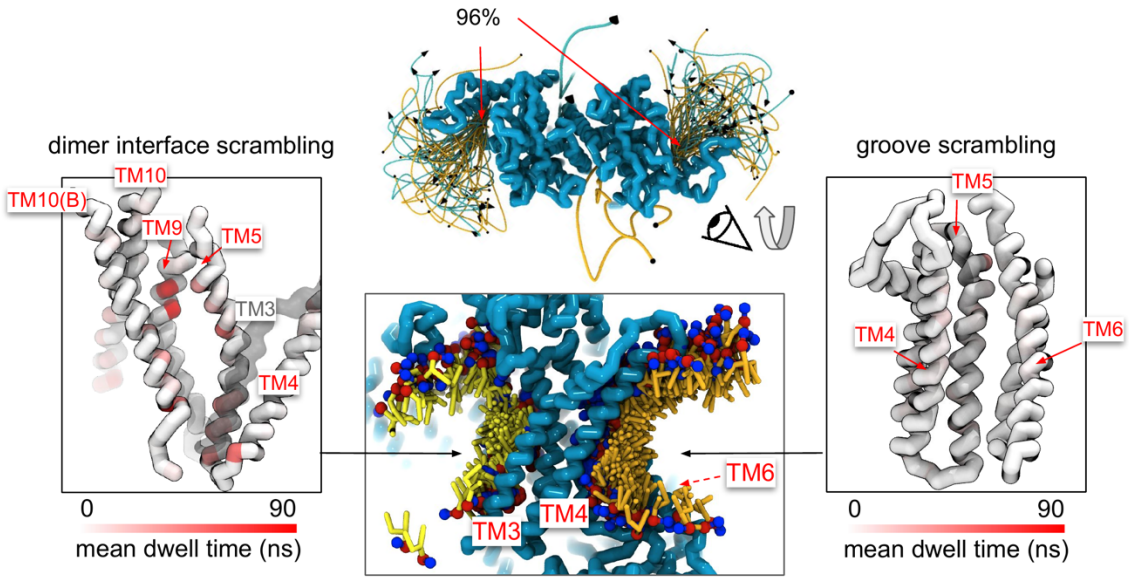

Figure 3.4 Lipid scrambling events and lipid-protein residue contact in the dimer interface and canonical TM4/6 groove.
Traces of all scrambling lipids in a TMEM16F (PDB ID 8B8J) simulation (center top). Lipids scramble from inner to outer leaflet are illustrated as cyan traces and from outer to inner leaflet as yellow traces. A cartoon depiction of two inward individual scrambling events along the TM4/6 groove (orange) and the dimer interface (yellow) with multiple snapshots over time (center bottom). Only headgroup, first and second tail beads are shown for clarity. Protein backbone colored by mean lipid headgroup interaction (dwell) time at the TMEM16F dimer interface (left) and TM4/6 groove (right).

Of the remaining six "out-of-the-groove" events one occurred across a closed TM4/6 groove of Ca^{2+} -bound TMEM16F (PDB ID 6P47). From all our observed scrambling events this is the only one that fits the "out-of-the-groove" model where scrambling is expected to take place where the membrane thins the most [104] (**Supplementary Information Figure 3.31*A***). Two events occurred along TM3 and TM4, which serve as a sort of channel the lipid headgroup, near the TM4/6 groove of an open nhTMEM16 (PDBID 4WIS) and predicted conductive state of TMEM16A (7ZK3*[8]) (**Supplementary Information Figure 3.31*B-C***). Lastly, two events occurred concurrently along TM6 and TM8 again near the hydrophilic groove of a Ca^{2+}-bound closed TMEM16F (PDB ID 8TAG) (**Supplementary Information Figure 3.31*D***). In each of these four "out-of-the-groove" events the scrambling lipid transverses with 2-4 water molecules around its headgroup (**Supplementary Information Figure 3.31**).

Discussion

In this work we have taken a systematic approach to quantify lipid scrambling by five TMEM16 family members and relate their scrambling competence to their structural characteristics and ability to thin the membrane. Currently there are two models for how TMEM16s transport lipids. In the first model the hydrophilic groove outlined by TM4 and TM6 dilates and allows lipids to insert their head groups into the hydrate cavity i.e. the "credit-card" [102] or "in-the-groove" model [104]. In the second model an open groove is not necessary, and the membrane deformation induced by the protein sufficiently lowers the energetic barrier to allow lipid crossing ~10^6-fold faster than spontaneous flipping [111][104]. This latter model offers a plausible explanation for scrambling by fungal TMEM16s with closed grooves or in the absence of Ca2+ and the lack of open-groove structures of Ca^{2+}-bound TMEM16F. Our results show that although there is a positive correlation between membrane thinning and scrambling (**Figure 3.2*C***), it is not the only basis for scrambling, at least within the detection limits of our CGMD simulations. For example, our simulations show that several TMEM16F structures (PDB IDs 6QPB, 6QP6, and 6QPC) and one TMEM16A structure (simulated PDB ID 7ZK3*[8]), thin the membrane more than the Ca^{2+}-bound afTMEM16 structure which has the second highest scrambling rate (**Figure 3.2*A***), but none of them

scramble lipids in the vicinity of the groove i.e. at the location of maximal thinning (**Table 3.2**, **Supplementary Information Figure 3.8-9**). None of these structures, however, have separated TM4 and TM6 by at least 4.7 Å. In fact, we *only* saw in-the-groove scrambling events when the groove was wider than 5 Å. We have thus determined that the second key feature that accounts for most scrambling events is the groove being wide enough to allow lipid headgroups to insert into the water-filled cavity.

It was unclear how the constitutively active TMEM16F F518H mutant (PDB ID 8B8J), which takes on a closed conformation in experiment, would scramble lipids. In our simulations we see it's TM4 and TM6 move 2.5 Å away from each other and measure water permeation and scrambling rates on par with open states of afTMEM16 and TMEM16K (**Figure 3.3*A***, **Supplementary Information Figure 3.16**). AAMD simulations of Ca^{2+} -bound WT TMEM16F [57], and a F518K/Y563K TMEM16F *in silico* mutant [60] were able to sample four scrambling events and single event respectively following spontaneous separation of TM4 and TM6 at the hydrophobic gate formed by F518, Y563 and I612. This gate was first identified by Le *et al.* using their own AAMD simulations of a TMEM16F homology modeled based on open nhTMEM16 [54] which sample 2 scrambling events. AAMD simulations of nhTMEM16 have also shown that the groove further dilates up to 2 Å compared to the starting X-ray coordinates [44][38]. AAMD simulations also predict that TMEM16A can dilate its ion pore, formed by TM3-8 when PIP_2, which is necessary for activation [75][49], [52], is bound to the Ca^{2+} -bound state and suggests that solved structures may be of desensitized or rundown states [50]. Our observed groove dilations for multiple TMEM16 structures (**Supplementary Information Figure 3.33**) are at least in line with protein dynamics in reported atomistic simulations. Conversely, we did fail to sample groove dilation and ion conduction for intermediate nhTMEM16 (6QM6) as predicted by AAMD [44] which may be due to the limited flexibility allowed by the Martini3 elastic network. Interestingly, our measured water permeation rates for intermediate nhTMEM16 and mutant TMEM16F F518H (8B8Q) are fairly similar (160 and 212 events/μs). It seems possible that a slight rearrangement of their hydrophobic

networks would allow ions but not lipids to pass, making his mutant TMEM16F conformation a specific ion-conductive intermediate.

PIP_2 also modulates TMEM16F activity [112] and at least in WT TMEM16F the presence of PIP_2 stabilizes a conformation with a large kink in TM6 that's associated with a significant membrane deformation profile in nanodiscs (PDB IDs 8TAG and 6P48 tested in this work) [72][71]. Cryo-EM structures of mutant TMEM16Fs also solved in the presence of PIP_2 take on very different conformations from the WT and do not have this large kink in TM6 (PDB IDs 8B8J, 8B8Q and 8B8K) [64]. To assess how the conformation of the protein itself contributes to scrambling we chose not to include PIP_2 in any of our simulations. Of these structures only the F518H mutant 8B8J, which thinned the membrane the most and sufficiently opened, scrambled lipids in the groove. It could be possible the PIP_2 interactions may enable some conformational change in TMEM16F to allow scrambling in the other structures as it did for TMEM16A. After all, TMEM16A is more closely related to TMEM16F than the other TMEM16 scramblases [73]. At least in the absence of PIP_2, we believe our work adds significant support to the idea that all TMEM16 Ca^{2+} -dependent scrambling activity in the groove, including TMEM16F's, relies on separation of TM4 and 6.

Of the scrambling competent TMEM16 structures the open groove nhTMEM16 (PDBID 4WIS) is the fastest with a scrambling rate lightly more than double that of other homologs (**Figure 3.2-3**). On average it thins the membrane more and has a wider groove than the other open nhTMEM16 (PDBID 6QM6), open afTMEM16, and simulated open TMEM16F (**Figure 3.3**). However, open TMEM16K (PDBID 5OC9) is slightly wider and thins the membrane more yet it has a ~3x lower scrambling rate and only 20% the number of water permeations (**Figure 3.3**). Unlike the fungal and TMEM16F and A homologs, TMEM16K resides in the endoplasmic reticulum (ER) membrane which is thinner than the plasma membrane [14], [35], [113]. Others have shown that TMEM16K scrambling rates can increase tenfold in thinner membranes which could make it the fastest scramblase we have tested here [14]. Lastly, the TMEM16F F518H mutant (PDBID 8B8J) thins the membrane more and has a larger flow of water

than nhTMEM16 but its TM4 and 6 are closer together and its scrambling rate is less than half (**Figure 3.2-3, Supplementary Information Figure 3.16**). This seems to violate the pattern amongst the scrambling competent structures that more membrane deformation and a wider groove translates to higher scrambling rates. Perhaps TMEM16F, or at least this constitutively active mutant, relies more heavily on membrane thinning than an open groove for scrambling and ion permeation. However, it is difficult to distinguish how much the thinning contributes to its scramblase activity given that groove also dilates. It will be interesting to see if any future experimental structures of WT TMEM16F have this unique bend in TM3 or a dilated groove similar to the other open TMEM16s. Lastly, even though in our simulations one of the predicted conductive states of TMEM16A open the groove enough to allow lipids to pass, they do not sufficiently thin the membrane enough to allow robust scrambling in the more constricted range (TM4-6 distances ~5-7 Å) of scrambling competent TMEM16s. Consistent with this the energy barrier for lipids in groove is up to 5.5-fold larger than for the scramblases (**Supplementary Information Figure 3.23*B***). Based on the trends we see for other TMEM16s it seems likely that if TMEM16A thinned the membrane more it would increase its scrambling rate but it is unclear if dilating the groove further we would have the same effect.

Lipids lining the TMEM16A groove

In simulations of our three predicted conductive states of TMEM16A (7ZK3*[6,8] and [10]) lipid headgroups insert into the lower and upper vestibule of the groove and thus partially line the ion conduction pathway. Surprisingly our simulation of the predicted TMEM16A conductive states by Jia & Chen [50] did have a fully-lipid lined groove, similar to how they form the proteolipidic pore in dual-function members (**Supplementary Information Figure 3.16**), however we did not observe any ion permeation events, so it is unclear how this functionally relates to its ion channel activity. Interestingly, recently published results show that the OSCA1.2 mechanosensitive ion channel, which has the same TMEM16 fold, forms a proteolipidic ion pore like the dual-function TMEM16 scramblases [114] and yet it does not scramble lipids [53]. One way to experimentally test if lipids are involved in ion permeation in

TMEM16A would be to measure ion uptake in liposomes of different lipid mixtures and observe any changes in selectivity.

"Out-of-the-groove" scrambling events

Experimental scrambling assays performed by different groups have reported basal level scramblase activity in the absence of Ca^{2+} for fungal and mammalian dual-function scramblases [13], [14], [15], [21], [42], [45], [59] but it's undetermined where closed-groove scrambling takes place on the protein [66]. In aggregate, we have observed 54 scrambling events that do not follow the credit-card model and would occur "out-of-the-groove" (Table 3.1). To our knowledge this the first reported prediction of TMEM16 scrambling outside of the hydrophilic groove and in closed-states and nearly all these events occur at the dimer interface. Although Li *et al* also reported scrambling events from closed TMEM16A, TMEM16K, and TMEM16F, but they did not specify where these events occurred [108]. We know from cryo-EM images that lipids intercalate between TM10 and 3 of fungal TMEM16s, TMEM16F and TMEM16K and there is also strong evidence that the PIP_2 lipid, a potent modulator for TMEM16A and TMEM16F activity [49], [50], [112][52], interacts with the protein in the dimer interface. However, there is no experimental evidence to support that lipids also scramble in the same location.

It is still curious that scrambling in the dimer interface only happens in the mammalian TMEM16s. Mammalian TMEM16s have a ~4-5 Å wider gap on average between TM10 and TM3 where lipids enter the interface than the open fungal TMEM16s, however, closed and intermediate states of nh and afTMEM16 and closed Ca2+-bound states of TMEM16A (excluding simulated states) also have gaps as wide as the mammalian structures but do not scramble lipids in the interface (**Supplementary Information Figure 3.32**). Additionally, there are several TMEM16F structures that do not have interface scrambling despite having the widest TM10-3 distances. We also see lipids in the dimer interface for all structures coming from both the inner and outer leaflets, even those where we don't see those lipids scramble, so accessibility to the interface does not seem to be a major determining factor (**Supplementary Information Figure 3.29**). We even see lipids "hop" over the TM10/3 constriction of

the open fungal structures to enter the interface. The ability to scramble lipids at the dimer interface may have a more nuanced explanation. We never observed lipids enter the interface from the lower leaflet in our atomistic nhTMEM16 simulations and thus the CG forcefield and representation may therefore be positively influencing the ability of lipids live in the dimer interface, although the AA aggregate simulation time was only 20% of our CG simulations time.

It is ambiguous whether the out-of-the-groove events we've observed reflect the same closed-groove scrambling activity seen in experimental assays [42], [45], [62], [66]. In 7:3 DOPC:DOPG liposomes Feng *et al.* [66] reports a ~7-15 fold increase in scrambling rate by Ca^{2+} -activated nhTMEM16 compared to apo. By applying the same fold increase to our own simulations, we would have expected to observe ~15-30 events for closed apo states of nhTMEM16 in a 9 μs simulation, however we observed no lipid scrambling. One reason for this discrepancy could be the difference in membrane composition (we used a pure DOPC bilayer) but more work is needed to clarify why. Our simulations of open TMEM16Fs however overestimate experimental estimates 10-100-fold [59] and likely overestimate closed groove scrambling as well but more experimental data is needed to make a full comparison. Interesting, our simulation of closed Ca^{2+}-bound TMEM16F (PDB 6P47) is the closest to the experimental estimate with its one out-of-the-groove event. Unlike the fungal scramblases and TMEM16F, our prediction for closed-groove scrambling by TMEM16K is within an order of magnitude of the experimental estimate in a POPC/POPG membrane [14] and thus dimer interface scrambling could explain its scrambling activity in the absence of an open groove. Finally, we saw that predicted conductive states of TMEM16A can also scramble, both in and out of the groove which does not fit with its experimental status as a non-scrambler. However, the scrambling rate we've predicted is so low, that it may not have been detectable in experiments [56], [107] and no one has tested for TMEM16A scrambling activity with the same dithionite-quenching liposome assay used for the dual-function scramblases. We are hopefully that our work will inspire new investigations into whether scramblase activity occurs at the dimer interface in a Ca^{2+}-independent manner as well as the role of lipids in TMEM16A ion conduction.

Methods

Starting structure selection

We chose to simulate 23 out of the 62 cryo-EM and X-ray TMEM16 homodimer structures available at the time (**Supplementary Information Table 3.1**). We chose not to include structures with greater than 4 Å global resolution, apart from TMEM16A 5OYG which is the only TMEM16A apo representative and structures with either large terminus truncations (PDB ID: 6BGI, PDB ID: 6BGJ and PDB ID: 8BC1) or largely unmodelled C-terminus domains (PDBID: 6QPI). We further narrowed our final set of structures by selecting the higher resolution of structures sharing similar backbone conformations and number of Ca2+ ions bound (i.e. PDB ID: 6QM9 and PDB ID: 6QM5 to PDB ID: 4WIS, PDB ID: 6OY3 to PDB ID: 6QMA, PDB ID: 6R65 to PDB ID: 5OC9, PDB ID: 7RX3, PDB ID: 7RXA, and PDB ID: 6DZ7 to PDB ID: 7RXB, PDB ID: 7RX2, PDB ID: 7RWJ and PDB ID: 6E0H, and PDB ID: 6E1O to PDB ID: 7RXG, PDB ID: 8B8M chain B to PDB ID: 8B8J, PDB ID: 8SUR, PDB ID: 8SUN, PDB ID: 8TAI and 8 PDB ID: 8TAL to PDB ID: 8TAG, PDB ID: 6P49 to PDBID: 6P48, PDB ID: 7B5C and PDB ID: 7B5E to PDB ID: 5OYB, PDB ID: 7B5D to PDBID: 5OYG). One structure from TMEM16A (PDB ID: 8QZC) and TMEM16F (PDB ID: 6P46) were excluded because they share a similar conformation to other structures of the same homolog but differ in the number of bound Ca2+ ions at the orthosteric site and by a slight elevation the TM6 C-terminus. It's unclear how the Ca2+ occupancy or subtle change to TM6 would alter the membrane around these closed TMEM16 structures, and perhaps worth exploring at a later point. We selected 5 nhTMEM16, 2 afTMEM16, 2 TMEM16K, and 11 TMEM16F dimers. We also include 2 cryo-EM structures of TMEM16A to see if the CG simulations can distinguish between scramblases and non-scrambles. We also chose to simulate a computationally predicted scrambling competent TMEM16F structure based on 6QP6 [57]which has a dilated groove similar to nhTMEM16, afTMEM16 and TMEM16K. We also simulated several computationally predicted conductive states of TMEM16A: one based on PDBID 5OYB [50]and three based on simulations 1PBC-bound 7ZK3 after removing 1PBC from the pore (unpublished work) which have

significant changes to TM3 and 4 compared to their starting experimentally determined structures. Several new nhTMEM16 structures were recently published by Feng et al. [66] but not released until after completion of our simulation work and therefore not included here.

General simulation details

All molecular dynamics simulations were performed with Gromacs (version 2020.6 [89]) and the Martini 3 coarse-grained (CG) force field (version 3.0.0 [106]). A 20 fs time step was used. Reaction-field electrostatics and Van der Waals potentials were cut-off at 1.1 nm [115]. As recommended by Kim *et al.* [116], the neighbor list was updated every 20 steps using the Verlet scheme with a 1.35 nm cut-off distance. Temperature was kept at 310 K using the velocity rescaling [117] thermostat (τ_T=1 ps). The pressure of the system was semiisotropically coupled to a 1 bar reference pressure by the Parrinello-Rahman [94] barostat (τ_P=12 ps, compressibility=3x10^{-4}).

Loop modeling

For each simulated structure missing loops with less than 16 residues were modeled using the loop building and refinement procedures MODELLER (version 10.2 [84]). Further details on which loops were included are in Supplementary Information Table 3.1. For each stretch of N missing residues 10xN models were generated. We then manually assessed the 10 lowest DOPE scoring predictions and selected the most biochemically reasonable model. Models were inserted symmetrically into the original experimental dimer structure except for PDBs 8BC0, 8TAG, and 5OC9 which were published as asymmetric structures.

Atomistic simulation details

See Chapter 2 Methods for system setup and simulation details for TMEM16A. An identical simulation protocol was used for our atomistic simulations of TMEM16F (PDB 6QP6) but used Gromacs version 2018.7. Missing residues 84-88, 143-206, 225-228, 428-444, 490-505, 588-590, 641-644, and 792-794 were modeled using MODELLER (version 10.2 [84]). Simulation details for TMEM16A performed by the Chen group are detailed in Jia & Chen 2021 [50]. Simulation details for atomistic

simulations of TMEM16F (PDB 6QP6) performed by the Weinstein group are detailed in Khelashvili et al 2022 [57]. Simulation details for atomistic simulations for nhTMEM16 are detailed in Bethel & Grabe 2016 [38].

Simulated TMEM16A structure selection

See Chapter 2 Methods for TMEM16A tIC and MSM analysis to cluster conformations captured by AAMD. We selected the medoids from three of these clusters (6, 8, and 10) which are more dilated that the starting structure and predicted to represent different conductive states. Note that this cluster numbering scheme is outdated and the medoids now correspond to clusters 7, 8, and 12 as presented in Chapter 2.

CG system setup

Setup of the CG simulation systems was automated in a python wrapper adapted from MemProtMD [37]. After preparing the atomistic structure using pdb2pqr [118], the script predicted protein orientation with respect to a membrane with memembed [119]. Then, martinize2 [120] was employed to build a Martini 3 CG protein model. Secondary structure elements were predicted by DSSP [121] and their inter- and intra-orientations within a 5-10 Å distance were constrained by an elastic network with a 500 kJ mol^{-1} nm^{-2} force constant (unless specified otherwise). CG Ca^{2+} ions (bead type "SD" in Martini 3) were inserted at their respective positions based on the experimental protein structure and connected to coordinating (<= 6 Å) ASP and/or GLU side chains by a harmonic bond with a 100 kJ mol^{-1} nm^{-2} force constant. A DOPC membrane was built around the CG protein structure using *insane* [122] in a solvated box of 220x220x180 $Å^3$, with 150 mM NaCl. Systems were charge-neutralized by adding Cl^- or Na^+ ions. For each system, energy minimization and a 2 ns NPT equilibration were performed. All systems were simulated for 10 μs in the production phase.

Density & maximum density path calculations

First, each protein subunit was individually aligned in x, y and z to their starting coordinates. Atomistic simulations were filtered for trajectory frames with T333-Y439 Cα distance >15 Å giving a

total of ~2085 ns of aggregate simulation time. Then the positions of all DOPC (POPC in atomistic simulations) headgroup (PC) beads was tracked overtime and binned in a 100x100x150 Å grid with 0.5 Å spacing centered on two residues near the membrane midplane on TM4 and 6 using a custom script that includes MDAnalysis methods[98], [123]. Density for water beads was calculated in the same way. Density in each cell was then averaged from each chain and for atomistic simulations averaged from all 8 independent simulations. One dimensional paths through the density were calculated by first selecting a single grid cell near the center of the box, totaling the density of all cells within a 4.7 Å cutoff and then repeating this step for each cell within a 9.4 Å of the first, until a maximum density total is identified. We then save the centroid of this final set of grid cells and repeat the search using this centroid, or node, as the starting position. This process continues until either the path length (sum of distances between subsequent nodes) reaches 80 Å for lipids (30 Å for water), no new nodes are found, or the next selected node causes the path to deviate sharply (<90°). We only included cells with densities >=0.0005 for lipids or >=0.002 for water for this calculation. Alternative paths were calculated by repeating the search but excluding cells used to define the nodes of the maximum density path. Path nodes where then interpolated using a B-spline representation [99] and final nodes were selected from this path to given 0.5 Å gaps between nodes for the lipid paths (4.7 Å for water).

Water permeation analysis

For each simulation the position of water CG beads within 9.4 Å search radius of the maximum density pathway were tracked overtime using a custom script that includes MDAnalysis methods [98], [123]. Water beads were assigned to path nodes if that fell within the bounds of a cylindrical disc (4.7 Å height, 9.4 Å diameter) centered on the bead with its face normal defined by the vector between the current and subsequent node. Permeation events were counted if a water bead left the search radius and its last bin assignment was in the opposite half of nodes as when it entered the path. The maximum density path was mirrored symmetrically on both subunits and each pathway was tracked independently. The first 1 μs was omitted from this analysis for equilibration. We also calculated the flux of water between nodes

by counting the net number of waters entering and leaving a cylinder (of the same proportions above) centered on each path node at each timestep (/ns). The residues chosen for measuring the distance between TM4 and TM6 were located within ~6 Å in z (1-2 α-helix turns) of the path node with the minimum net flux of water. The residues used for each homolog were as follows: 327-339 and 430-452 for nhTMEM16, 319-331 and 426-438 for afTMEM16, 365-377 and 434-446 for TMEM16K, 512-424 and 613-625 for TMEM16F, and 541-553 and 635-647 for TMEM16A.

Scrambling analysis

Lipid scrambling was analyzed as described by Li et al [103]. For every simulation frame (1 ns sampling rate), the angle between each individual DOPC lipid and the z-axis was calculated using the average of the vectors between the choline (NC3) bead and the two last tail beads (C4A and C4B), see Supplementary Information Figure 3.5*A*. We applied a 100 ns running average to denoise the angle traces. Lipids that reside in the upper leaflet are characterized by a 150° angle, and lipids in the lower leaflet have a 30° angle. Scrambling events were counted when a lipid from the upper leaflet passed the lower threshold at 35° or, *vice versa*, when a lipid from the lower leaflet passed the upper threshold at 145° (see **Supplementary Information Figure 3.5*A***). These settings are more stringent than the thresholds used by Li et al (55° and 125°, respectively) to prevent falsely counted partial transitions.

The first 1 μs was omitted from the scrambling analysis for equilibration. Similar to the protocol by Li et al [103], a 1 μs block averaging was applied to the remaining 9 μs to obtain averages and standard deviations for the scrambling rates. This was also applied for the water permeation rates.

Quantification of membrane deformations

First, using Gromacs (gmx trjconv), MD trajectories were aligned in the xy-plane such that the longest principal axis defined by the initial positions of transmembrane helices 7 and 8 aligned to the global y axis. Average membrane surfaces were calculated from the aligned MD trajectories as outlined previously [38] using a custom python script based on MDAnalysis [98] and SciPy [99]. The positions of each lipid's glycerol beads (GL1 and GL2) were interpolated to a rectilinear grid with 1 Å spacing.

Averaging over all time frames (again, discarding the first 1 μs for equilibration) yielded a representative upper and lower leaflet surface. The difference between the two surfaces spans the hydrophobic membrane interior and was used as a measure for the membrane thickness along the bilayer normal (**Fig. 2-3**). Crucially, in the case of lipid scrambling simulations like the ones described here, lipids were assigned to the upper/lower leaflet separately for every time frame.

Protein-lipid contact and dwell time analysis

Using the full 10 μs simulation where each protein subunit was individually aligned in x, y, and z, we analyzed protein-lipid interactions by measuring distances between the protein's outermost side chain bead (except for glycine, which only has backbone bead) and the lipid's choline (NC3) or phosphate (PO4) bead for every nanosecond (saving rate) using custom scripts with Scipy methods [99]. Contacts were defined as distances below 7 Å. Contact frequency was calculated as the fraction of simulation frame where a contact occurred, averaged over two monomers. Dwell time was measured as the duration of a consecutive contact, allowing breaks up to 6 ns to account for transient fluctuations of lipid configuration. For each residue, we selected either the choline or phosphate bead based on which yielded the higher average dwell time. To visualize the result, we used averaged dwell time of top 50% longest dwelling events at each residue to generate a color-coded representation of the protein structure.

PMF calculations

To calculate the PMFs for each pathway we assigned each cell from the average PC density to its closest respective mean-lipid pathway node. We then summed the density of all cells assigned to a given node and divided by the total volume of those cells. We used the following equation to convert the density values to energies or PMF.

(1) $$E = \frac{-R*T*\ln(\rho/\rho_0)}{1000}$$

0, is the density of PC in a 10 μs protein-free membrane simulation, R is the gas constant.

4. SHEDDING LIGHT THROUGH AN OPEN DOOR: A CONNECTION BETWEEN LIGAND INTERACTION DYNAMICS AND PROTEIN REARRANGEMENTS IN A FLUOROGEN ACTIVATING COMPLEX

Introduction

In 2013 the Bruchez lab identified a ternary complex composed of a single malachite green (MG) dye analog bound to a light chain antibody dimer (dL5*) that exhibits high fluorescent quantum yield (QY) when exposed to an activating wavelength of light [124]. Later in 2015 they developed an iodinated MG analog that when in complex with a mutant dL5* protein produces excited singlet oxygen ($^1O_2$) when exposed to near infrared wavelengths[125]. This fluorogen-activating complex could be useful for imaging and if developed into a medical treatment as a highly localized light-based therapy for tumor cell ablation. Since this published work they have synthesized a series of iodinated MG analogs with a range of $^1O_2$ QYs. All the dyes possess a common core consisting of three linked aromatic rings, but they differ in a tail region that attaches to one of the rings (the oxygen-carrying ring) and protrudes from the protein into solution. Since the core of all the dyes is conserved and it is this region of the molecule that likely gives rise to its spectroscopic properties, it is unclear why changes in the tails impact singlet oxygen generation. To begin to address this problem, I employed AAMD simulations to determine the properties of the dyes in solution and in the protein. The simulations reveal that the tails of some the dyes have extensive, long-lived interactions with residues on the protein surface. Moreover, in the presence of the antibody dimer, correlations emerge showing that differences in the tails impact the conformation of the aromatic rings, which may tune the $^1O_2$ QY properties of the dyes. Additionally, dye molecules whose tails have extensive long-lived (i.e. less dynamic) contact with the protein can induce a conformational change that widens the dimer interface at the opposite end from where the tail protrudes increasing water and oxygen access to the core, in a process we term “cracking”. We only observe cracking for three of the dye analogs which all produce 2-5x more $^1O_2$ than the remaining fourth analog. Taken together this shows that cracking and dye molecule dynamics have a positive effect on $^1O_2$ QY.

Results

I simulated multiple replicas of four dye analogs (MG2I-Ester-R, -Ester, -Amide, and -Sulfonate) free in solution or bound to the mutant dL5* in zero, low (0.01 mM), or high (0.1 mM) molecular oxygen (MO), and additional simulations tested two more dyes (MG2P and MG2I-OH) under in 0 mM MO. Each run lasted 1 to 3 μs, and the aggregate data set is 217 μs (**Supplemental Information Data Table 4.1**). The quantum mechanical (QM) parameterization of the dye properties used in the MD simulations (**Supplemental Information Figure 4.1**) reproduces mean geometric angles observed in the bound MG-2p structure (PDB 4K3H) [124] as well as minimum energy configurations in solution predicted from QM geometry optimization (**Supplemental Information Figure 4.2**) more faithfully than the default Antechamber parameters. We repeated the same parameterization process for the iodinated dye (**Supplemental Information Figure 4.3**). All iodinated dye analogs better sample the X-ray and QM optimized structures values for the oxygen-carrying ring rotation when simulated in solution than when bound to dL5*, which experiences a near 15° reduction in the rotation angle (**Supplemental Information Figure 4.4**). This predicated shift in angle is likely due to the increased volume introduced by the two iodine atoms near the binding site entrance and steric clashes with nearby residues. Given the dramatic increase in fluorescent QY observed upon binding [124], I expected a reduction in dye ring motility when bound which reduces the likelihood of non-radiative decay of excited state energy through thermal vibrations. This is similar to the mechanism of protein-induced fluorescence enhancement explored experimentally with Cy3 dyes which have longer fluorescence lifetimes when constrained by a protein or placed in a more viscous medium [126], [127]. To this end, puckering is more highly constrained in the presence of the protein (data not shown), and the autocorrelation of the ring rotation shows a slow exponential decay for all dyes on the 100s of nanosecond timescale while in solution the motion becomes completely uncorrelated in less than a nanosecond (**Figure 4.1A**). Thus, the motion of the rings is correlated with the protein dynamics. We also compared the angular velocity of dye ring rotation and observed a higher frequency of low magnitude velocities when the dye was bound to the protein than in

solution (**Figure 4.1B**). Together these simulations demonstrate that the protein has a profound impact on dye dynamics and is consistent with our idea that the protein suppresses the nonradiative deactivation mechanism by dampening molecular motion.

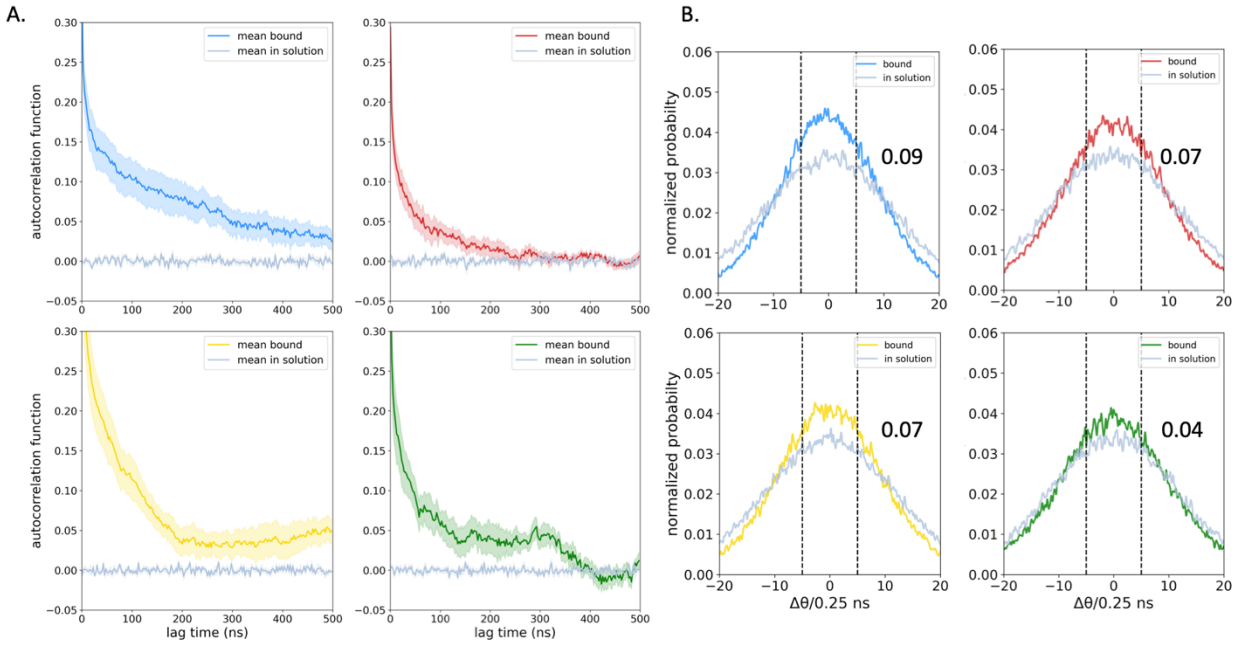

Figure 4.1 Autocorrelation functions of the oxygen-carrying aromatic ring.
(A) Each plot shows a mean of the autocorrelation function of O-carrying ring rotation with lag times up to 500 ns from simulations of MG2I-Ester-R (blue), -Ester (red), -Amide (yellow), and -Sulfonate (green) bound to dL5* and in solution (steel blue). Shaded regions represent the standard error of the mean at each lag time. (B) Probability distributions of angular velocity of the O-carrying ring from all simulations of MG2I-Ester-R (blue), -Ester (red), -Amide (yellow), and -Sulfonate (green) bound to dL5* and in solution (steel blue). Inlaid value displays the middle Reimann sum of the space between bound an in-solution curves from -5 to 5 $\Delta\theta$/0.25 ns (dotted black vertical lines).

Tail-protein contact lifetime correlates with higher $^1O_2$ QY

Next, I wanted to determine if there was any coupling between the tail, the core, and the protein and whether the degree of coupling changed with different tails. To do this, I selected 17 system and dye-specific features and computed the covariance matrix between all of these features for each dye in solution and bound to dL5* (**Figure 4.2B, C**). There was a near perfect correlation (>0.9) between the extent of tail contact with the protein ('tail-protein contact') and the rotation angle of the tail ('linker rotation') because as the tail moved to extend into solution or touch different portions of the protein

surface it required a concomitant change in the tail linker angle (**Figure 4.2B**). The next strongest covariance signals emerged between these two tail features and the rotation ('N-ring rot.') and puckering ('N-ring pucker') of the two N-carrying rings in the bound state (**Figure 4.2B**), which are completely absent in solution (**Figure 4.2C**). This revealed that the protein biases particular conformations of the dye while also establishing a direct connection between the behavior of the tail and that of the aromatic core. Importantly, it shows that the conformations of the aromatic rings are highly connected to how well the tail touches the protein.

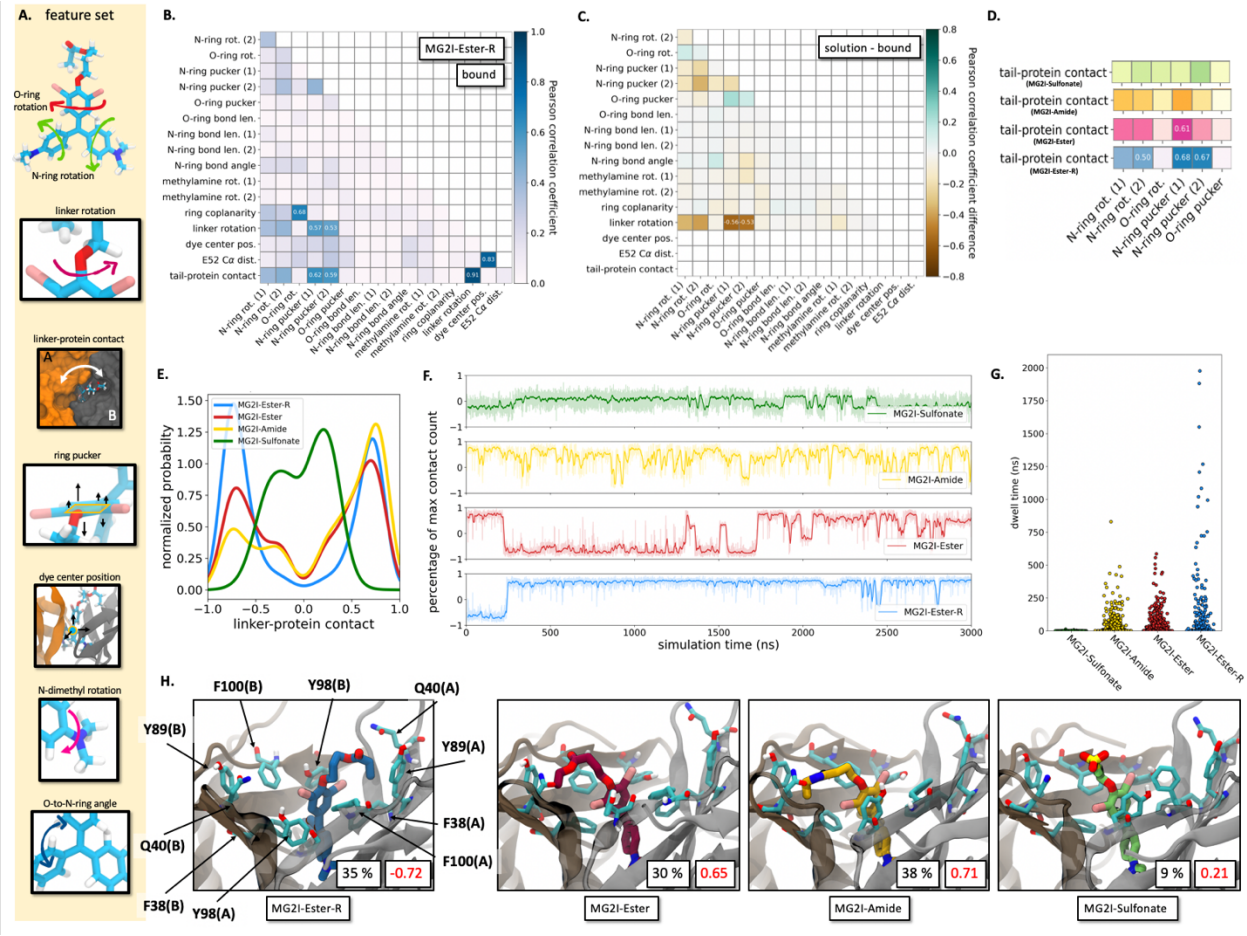

Figure 4.2 Dye analog feature covariances indicate moderate correlation between linker dynamics and aromatic ring behavior.
(A) Subset of the 17 features used to examine correlations in the system. (B) Covariance matrix of all 17 features extracted from the bound MG2I-Ester-R simulations. (C) Covariance matrix of dye-only features extracted from free and bound MG2I-Ester-R simulations. Matrix shows the difference between covariance coefficients from solution and bound. (D) Covariance of aromatic ring features with tail-protein contact with |contact values| >= 0.8. (*Figure caption continued on the next page.*)

(*Figure caption continued from the previous page.*) (E) Linker-protein contact distributions. (F) Time series of the tail contact against subunit A (-1), no protein contact (0), and subunit B (+1). (G) Linker-protein contact dwell times. (I) Cluster analysis performed on the tail for dyes bound to dL5* showing the medoid from the top cluster for MG2I-Ester-R, -Ester, -Amide, and -Sulfonate. Inlaid values show the percentage of input simulation frames within the cluster (black) and the mean linker-protein contact value (red).

The strength of the correlation between tail-protein contact and aromatic ring properties matches the degree of $^1O_2$ QY measured in the experiments, as MG2I-Ester-R exhibits the highest correlation, and it is the strongest singlet oxygen producer while MG2I-Sulfonate has the weakest signal, and it is the lowest producer (**Figure 4.2D**). In fact, MG2I-Sulfonate spends most of the simulation with its tail fully solvated making little contact with either subunit, while the other dyes make progressively greater contact with the protein, and MG2I-Ester-R has the tightest bimodal distributions each with high protein contact (**Figure 4.2E**). The dynamics of the tails again mirrored the experimental trend as reflected in how frequently tails switched from contacting subunit A (-1) to contacting subunit B (+1), where more frequent switching correlated with lower $^1O_2$ QY (**Figure 4.2F**) as quantified in the tail dwell times (**Figure 4.2G**). Due to the symmetry of the dL5* dimer and flexibility of the tails, I expected symmetric protein-contact distributions with near equal heights on either subunit; however, the slow switching dynamics of the higher QY dyes may have led to insufficient sampling. Lastly, I carried out clustering analysis on the tails to determine if they adopt a well-defined pose against the protein and identify residues responsible for stabilizing these poses (**Figure 4.2*H***). For MG2I-Ester-R, -Ester, and -Amide 60-80 % of the snapshots fall into 2 symmetric clusters, one against subunit A and the other against subunit B, that show extensive contact with large hydrophobic residues F38, Y89, Y98, and F100, and more distant contact with Q40. Meanwhile, using the same cluster criteria, the largest cluster for MG2I-Sulfonate comprised only ~8 % of the snapshots (including both symmetric poses) and the end of the tail was fully solvated. The hydrophobic nature of the adjacent protein residues coupled with the formal -1 charge on the sulfonate tail likely explains the dynamics and lack of protein contact for this dye compared to the others. The dominant pose identified from clustering tail coordinates had shorter distances between

the tail carbonyl and Y89 and F100 aromatic rings for MG2I-Amide and MG2I-Ester. This suggested that the shorter interaction lifetimes of these dyes may be due to clashes between these atoms even in the most energetically favorable positions of their tails (**Supplementary Information Figure 4.5**).

Dyes with high $^1O_2$ QY induce a conformational change that increases water access to the aromatic core

Multiple replicas of bound MG2I-Ester-R, -Ester, and -Amide captured spontaneous transitions of dL5* away from the conformation observed in the X-ray structure for tens to hundreds of nanoseconds. These transitions involved a rigid body rotation of one subunit against the other with a pivot point centered on the dye aromatic rings (**Figure 4.3*A*** and ***B***). In a movement we term "cracking", salt-bridging residues E52 and K55 and hydrophobic pairs Y34 and Y57 that initially form intersubunit contacts across the back side (opposite the tail-protein interaction site) of the binding cavity separate increasing access to the N-carrying rings (**Figure 4.3*A*** and ***B***).

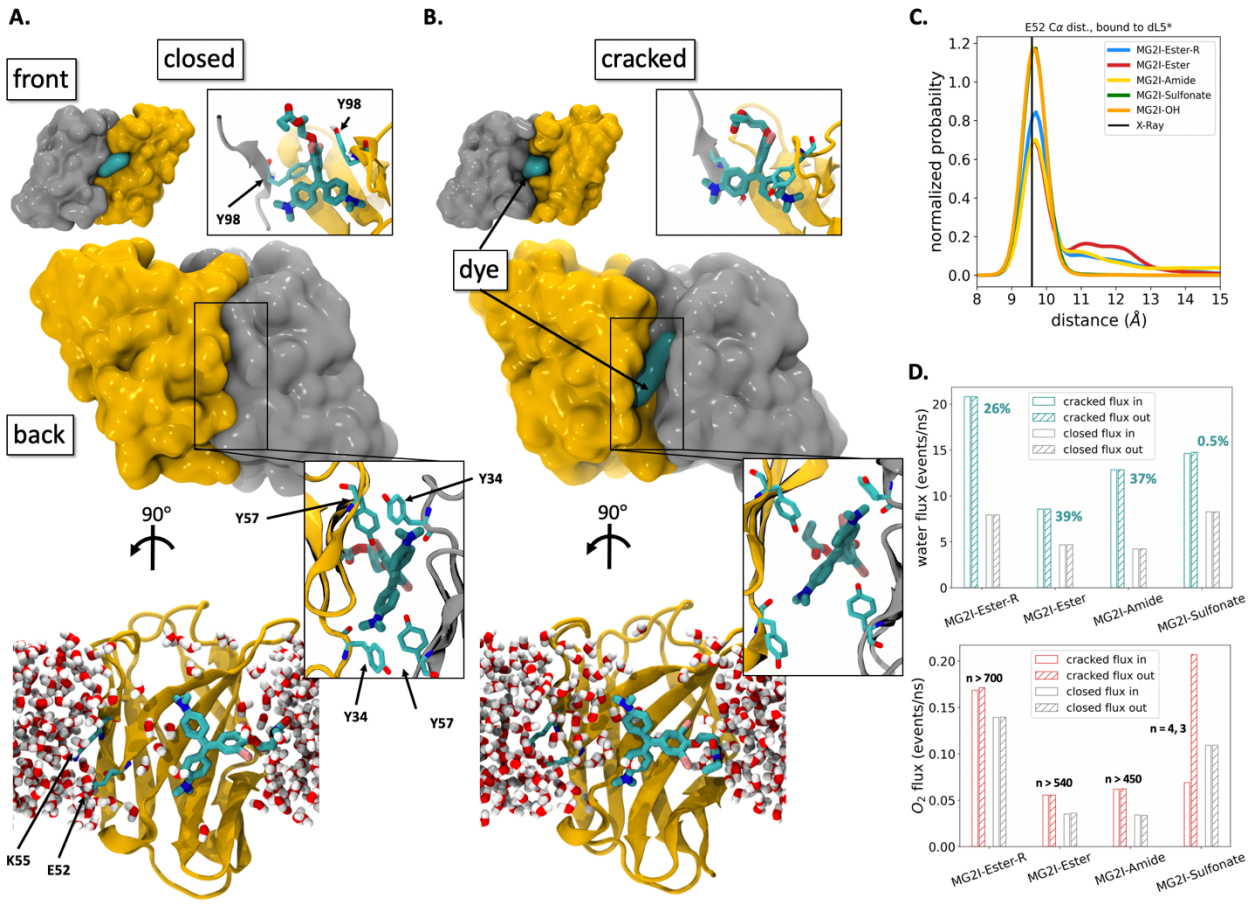

Figure 4.3 Tails with more stable interactions with protein residues induce dL5* subunit rearrangement and associated solvent placement near the bound dye molecule.
Images of MG2I-Amide dL5* in in the (A) closed and (B) cracked conformations from MD simulations. Each subunit is colored either yellow or grey and MG2I-Ester-R in cyan. (top left) a head-on view (front) of the bound dye tail. (top right) a closeup of the dye and Y98 residues. (middle) back view of the dimer and closeup of the dimer interface. (bottom) water around the back and front side of the dimer interface, one subunit is not shown. (C) Ca – Ca distance distributions between residues E52 from each subunit. (D) Solvent flux values in and out of the back dimer interface in cracked and closed conformations. Flux values for MO only shown for simulation on 0.1 M MO. Inlaid values for the top plot display the percentage of simulation data in the cracked state. Inlaid values in the bottom plot display the number inward and outward of MO flux events observed in the cracked states.

I used the E52 Cα - Cα distance to quantify this change (**Figure 4.3C**), and it is apparent from the distance distributions and time traces that MG2I-Sulfonate does not induce this conformational change. To test the hypothesis that tail contact to the protein drives the change, I simulated a truncated MG2I analog, MG2I-OH, bound to dL5*. Like MG2I-Sulfonate, this tailless analog failed to induce cracking, suggesting that tail-protein contact is in fact crucial. Lastly, I observed instantaneous differences

in solvent accessibility to the back of the dimer interface and 2 dye N-carrying rings (**Figure 4.3*A*, *B***). Three dimensional densities of water oxygens around the dye molecule, collapsed onto two dimensions at a time, reveal a change in the number of waters in the newly-opened portion of the binding site (**Figure 4.4**).

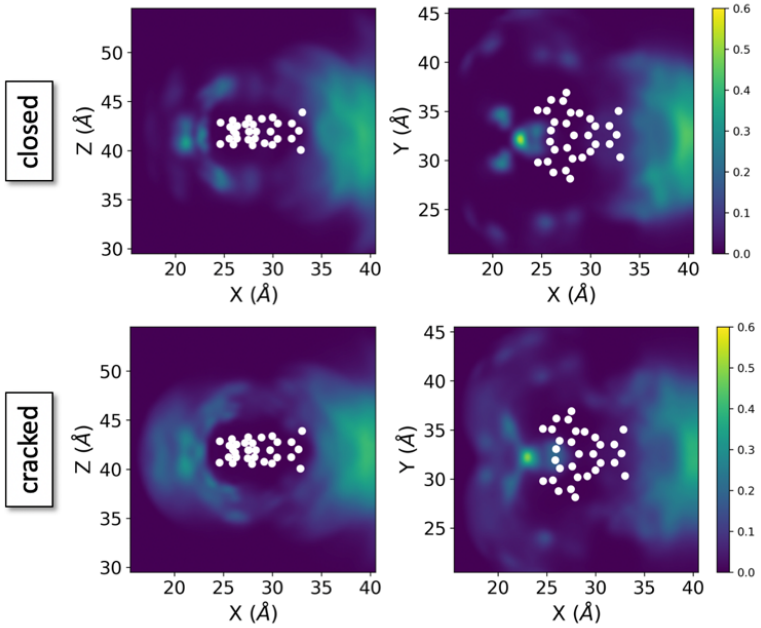

Figure 4.4 Changes in solvent density due to dL5* conformation change.
Density from water oxygens collected from all simulations of MG2I-Ester-R, -Ester, -Amide, and –Sulfonate when in closed (top) and cracked (bottom) dL5* conformations collapsed into 2D. White dots represent positions of dye carbon atoms.

I also saw an increase in the flux of water molecules into the same area of the back dimer interface while the protein was cracked (**Figure 4.3D**). I observed the highest flux values for MG2I-Ester-R, which is surprising given that MG2I-Ester and -Amide spend more time in the cracked states. One reason for this could be that I defined the cracked state using only the E52 Cα – Cα distance and this single feature may not accurately define the cracked state leading to variability in the data contributing to the flux calculations. I saw a similar trend on MO flux for MG2I-Ester-R, -Ester, -Amide and -Sulfonate, but we likely need more sampling for MG2I-Sulfonate to better determine if cracking changes MO accessibility to the aromatic rings. In brief, these MD simulations predict that the observed cracking

behavior brings more solvent into contact with the dye while also increasing the exchange rate of molecular oxygen already in contact with the dye, thus enhancing singlet oxygen generation. This cracking also changes the relative position of the dye in the dimer interface. I observed a new alignment dye N-carrying aromatic rings with the Y98 which formed displaced aryl stacks in in the cracked conformation (**Figure 4.3*A,B***). Changes in the emission properties, including increased fluorescence, of π-stacked conjugated aromatic molecules, have been noted in the field [128]·[129]. However, additional quantum mechanical calculations and experiments would be necessary to understand the impact this π-stacking has on the electronic nature of the bound MG analogs.

Discussion

Taken together my results show that cracking and dye molecule dynamics have a positive effect on $^1O_2$ QY. At the quantum level we theorize that the excited triplet state of the dye transfers its energy directly to a nearby ground state molecular oxygen thus exciting it to its singlet state. We hypothesize that cracking facilities longer-lived interactions between oxygen and the excited dye to allow for the energy exchange to occur. Originally two sequences were selected from the original yeast expression as good fluorogen-activating proteins for the un-iodinated malachite green (MG): dL5* and dL5**[1]. The only difference between the two sequences is that dL5** has a E52D mutation located near the dimer interface closest to the dimethylamidated MG rings and has a ~6-fold higher affinity for MG and therefore was used in experiments to measure the ability of iodinated-MG analogs to produce singlet oxygen. In simulations this cracking rearrangement was only observed for analogs with tails that made long lived contact with the binding site entrance. Indeed, when we simulated a bound analog where we replaced the tail with a hydroxyl group, no cracking occurred. How this protein contact with the tail is related to the conformational change remains unclear, but I speculate the effect may be allosteric. Given that E52 faces its c2-symmetric pair across the dimer interface (**Fig. 2A**) and that they move away from each other during cracking events, we hypothesized their charge repulsion drives this cracking behavior. This cracking behavior could explain the lower affinity of dL5* for MG. To test this hypothesis, I made an *in*

silico E52D dL5* mutant and ran simulations bound to two different analogs: MG2I-Sulfonate and MG2I-Ester. Only the latter sampled cracking in dL5* WT. I assume that the E52D mutation does not strongly change the structure of the individual subunits or overall location of the dimer interface. I found that the E52D mutation did not sample any cracking events in 9 μs of total simulation time with either dye analog bound (**Supplemental Information Fig. 4.6**). Comparison to the E52 (dL5*) simulations shows that E52 can take on more extended conformations into the dimer interface to interact with a nearby K55 on the same subunit and appears overall more dynamic than E52D. It is still undetermined how the protein tail interactions relate to cracking, but it is clear that E52 is critical for this behavior. I predict that a D52E mutant will have increased as cracking increases solvent accessibility to the dye in our simulations or that proteins bound to MG2I-Ester-R,-Ester, and -Amide will then have nearly identical SOG QY as they all sample cracking to a similar extent.

Methods

All simulations of the complex were initiated from the 2.45 Å X-Ray structure (PDB 4K3H), which contains an un-iodinated MG analog bound to wild type dL5*. Tail coordinates for each analog were generated with OpenBabel [130], and the iodine atoms were added with PyMOL [131]. When preparing the bound systems, we assumed that the core iodinated dyes bind in an identical manner to the non-iodinated MG. I manually refined their tail coordinates by rotating the dihedral angle linking the tail to the aromatic core to minimize clashes with the protein also in PyMOL[131]. Parameters for dye molecules were generated using Antechamber, a program within Amber that attempts to assign parameters based on similar chemistry pre-defined in the General Amber Forcefield (GAFF)[132], [133]. Additionally, special care was taken to parameterize the aromatic portion of each dye, including dihedral bond angles and Lennard-Jones constants for aromatic hydrogens by fitting single point quantum mechanical energies calculated to adjusted Amber potentials. Quantum mechanical calculations were done using Gaussian09 with the b3lyp Density Functional Theory model with the def2tzvp basis set for iodine atoms and 6-31G* basis set for all other atoms [134]. Given the electronic resonance of the aromatic rings we also enforced

symmetric parameter assignment across the molecule. I made the *in silico* E52D mutant protein, known as dL5** [124], using PyMOL 'Mutagenesis Wizard' [131]. I used Antechamber again to assign protein parameters from the Amber 14SB forcefield [134] [135] [136]. Waters were modeled using the Amber TIP3P water model and molecular oxygen parameters were assigned from the GAFF. All simulations, including minimization and equilibration, were performed with the Amber18 engine. Electrostatic interactions were described using the Particle Mesh Ewald (PME) algorithm [137] with a cutoff distance of 8 Å. Van der Waals interactions were also cutoff at 8 Å. We used a 2 fs MD timestep. I used Langevin dynamics with collision frequency of 5 ps^{-1} to maintain 300.0 K temperature and the MC barostat [138][139] to maintain an isotropic pressure of 1 bar. I wrote custom analysis scripts to quantify protein and dye features and measure solvent density. These scripts used tools from MDAnalysis [98], [123].

5. REFERENCES

[1] F. Khan *et al.*, "Membrane potential accelerates sugar uptake by stabilizing the outward facing conformation of the Na/glucose symporter vSGLT," *Nat Commun*, vol. 14, no. 1, p. 7511, Nov. 2023, doi: 10.1038/s41467-023-43119-z.

[2] M. Siggel, R. M. M. Bhaskara, and G. Hummer, "Phospholipid Scramblases Remodel the Shape of Asymmetric Membranes," *J Phys Chem Lett*, vol. 10, no. 20, pp. 6351–6354, Oct. 2019, doi: 10.1021/acs.jpclett.9b02531.

[3] J. Suzuki, T. Fujii, T. Imao, K. Ishihara, H. Kuba, and S. Nagata, "Calcium-dependent Phospholipid Scramblase Activity of TMEM16 Protein Family Members," *Journal of Biological Chemistry*, vol. 288, no. 19, pp. 13305–13316, May 2013, doi: 10.1074/jbc.M113.457937.

[4] E. Di Zanni, A. Gradogna, J. Scholz-Starke, and A. Boccaccio, "Gain of function of TMEM16E/ANO5 scrambling activity caused by a mutation associated with gnathodiaphyseal dysplasia," *Cellular and Molecular Life Sciences*, vol. 75, no. 9, pp. 1657–1670, May 2018, doi: 10.1007/s00018-017-2704-9.

[5] H. Kim *et al.*, "Anoctamin 9/TMEM16J is a cation channel activated by cAMP/PKA signal," *Cell Calcium*, vol. 71, pp. 75–85, May 2018, doi: 10.1016/j.ceca.2017.12.003.

[6] Y. D. Yang *et al.*, "TMEM16A confers receptor-activated calcium-dependent chloride conductance," *Nature*, 2008, doi: 10.1038/nature07313.

[7] A. Caputo *et al.*, "TMEM16A, a membrane protein associated with calcium-dependent chloride channel activity," *Science (1979)*, 2008, doi: 10.1126/science.1163518.

[8] H. Stöhr *et al.*, "TMEM16B, A Novel Protein with Calcium-Dependent Chloride Channel Activity, Associates with a Presynaptic Protein Complex in Photoreceptor Terminals," *The Journal of Neuroscience*, vol. 29, no. 21, pp. 6809–6818, May 2009, doi: 10.1523/JNEUROSCI.5546-08.2009.

[9] B. C. Schroeder, T. Cheng, Y. N. Jan, and L. Y. Jan, “Expression Cloning of TMEM16A as a Calcium-Activated Chloride Channel Subunit,” *Cell*, vol. 134, no. 6, pp. 1019–1029, Sep. 2008, doi: 10.1016/j.cell.2008.09.003.

[10] H. Yang *et al.*, “TMEM16F forms a Ca2+-activated cation channel required for lipid scrambling in platelets during blood coagulation.,” *Cell*, 2012, doi: 10.1016/j.cell.2012.07.036.

[11] J. R. Martins, D. Faria, P. Kongsuphol, B. Reisch, R. Schreiber, and K. Kunzelmann, “Anoctamin 6 is an essential component of the outwardly rectifying chloride channel,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 44, pp. 18168–18172, Nov. 2011, doi: 10.1073/pnas.1108094108.

[12] J. Suzuki, M. Umeda, P. J. Sims, and S. Nagata, “Calcium-dependent phospholipid scrambling by TMEM16F,” *Nature*, 2010, doi: 10.1038/nature09583.

[13] M. Malvezzi *et al.*, “Ca2+-dependent phospholipid scrambling by a reconstituted TMEM16 ion channel,” *Nat Commun*, vol. 4, no. 1, p. 2367, Sep. 2013, doi: 10.1038/ncomms3367.

[14] S. R. Bushell *et al.*, “The structural basis of lipid scrambling and inactivation in the endoplasmic reticulum scramblase TMEM16K,” *Nat Commun*, vol. 10, no. 1, p. 3956, Dec. 2019, doi: 10.1038/s41467-019-11753-1.

[15] B. C. Lee, A. K. Menon, and A. Accardi, “The nhTMEM16 Scramblase Is Also a Nonselective Ion Channel,” *Biophys J*, 2016, doi: 10.1016/j.bpj.2016.09.032.

[16] P. Scudieri *et al.*, “Ion channel and lipid scramblase activity associated with expression of TMEM16F/ANO6 isoforms,” *J Physiol*, vol. 593, no. 17, pp. 3829–3848, Sep. 2015, doi: 10.1113/JP270691.

[17] M. E. Falzone *et al.*, “Structural basis of Ca2+-dependent activation and lipid transport by a TMEM16 scramblase,” *Elife*, 2019, doi: 10.7554/eLife.43229.

[18] V. M. Milenkovic, M. Brockmann, H. Stöhr, B. H. Weber, and O. Strauss, “Evolution and functional divergence of the anoctamin family of membrane proteins,” *BMC Evol Biol*, vol. 10, no. 1, p. 319, 2010, doi: 10.1186/1471-2148-10-319.

[19] H. Terashima, A. Picollo, and A. Accardi, “Purified TMEM16A is sufficient to form Ca^{2+}-activated Cl^- channels,” *Proceedings of the National Academy of Sciences*, vol. 110, no. 48, pp. 19354–19359, Nov. 2013, doi: 10.1073/pnas.1312014110.

[20] K. Yu, C. Duran, Z. Qu, Y.-Y. Cui, and H. C. Hartzell, “Explaining Calcium-Dependent Gating of Anoctamin-1 Chloride Channels Requires a Revised Topology,” *Circ Res*, vol. 110, no. 7, pp. 990–999, Mar. 2012, doi: 10.1161/CIRCRESAHA.112.264440.

[21] J. D. Brunner, N. K. Lim, S. Schenck, A. Duerst, and R. Dutzler, “X-ray structure of a calcium-activated TMEM16 lipid scramblase,” *Nature*, 2014, doi: 10.1038/nature13984.

[22] H. C. Hartzell, K. Yu, Q. Xiao, L. Chien, and Z. Qu, “Anoctamin/TMEM16 family members are Ca^{2+}-activated Cl^- channels,” *J Physiol*, vol. 587, no. 10, pp. 2127–2139, May 2009, doi: 10.1113/jphysiol.2008.163709.

[23] C. Alvadia, N. K. Lim, V. C. Mosina, G. T. Oostergetel, R. Dutzler, and C. Paulino, “Cryo-EM structures and functional characterization of the murine lipid scramblase TMEM16F,” *Elife*, 2019, doi: 10.7554/eLife.44365.

[24] E. Castoldi, P. W. Collins, P. L. Williamson, and E. M. Bevers, “Compound heterozygosity for 2 novel TMEM16F mutations in a patient with Scott syndrome,” 2011. doi: 10.1182/blood-2011-01-332502.

[25] P. Boisseau *et al.*, “A new mutation of <*scp*>*ANO*</*scp*> 6 in two familial cases of Scott syndrome,” *Br J Haematol*, vol. 180, no. 5, pp. 750–752, Mar. 2018, doi: 10.1111/bjh.14439.

[26] M. B. Brooks, J. L. Catalfamo, R. MacNguyen, D. Tim, S. Fancher, and J. A. McCardle, “A TMEM16F point mutation causes an absence of canine platelet TMEM16F and ineffective

activation and death-induced phospholipid scrambling," *Journal of Thrombosis and Haemostasis*, vol. 13, no. 12, pp. 2240–2252, Dec. 2015, doi: 10.1111/jth.13157.

[27] S. Tsutsumi *et al.*, "The Novel Gene Encoding a Putative Transmembrane Protein Is Mutated in Gnathodiaphyseal Dysplasia (GDD)," *The American Journal of Human Genetics*, vol. 74, no. 6, pp. 1255–1261, Jun. 2004, doi: 10.1086/421527.

[28] I. Cabrita *et al.*, "TMEM16A mediates mucus production in human airway epithelial cells," *Am J Respir Cell Mol Biol*, 2021, doi: 10.1165/rcmb.2019-0442OC.

[29] F. Huang *et al.*, "Calcium-activated chloride channel TMEM16A modulates mucin secretion and airway smooth muscle contraction," *Proc Natl Acad Sci U S A*, 2012, doi: 10.1073/pnas.1214596109.

[30] D. A. Griffin *et al.*, "Defective membrane fusion and repair in *Anoctamin5* -deficient muscular dystrophy," *Hum Mol Genet*, vol. 25, no. 10, pp. 1900–1911, May 2016, doi: 10.1093/hmg/ddw063.

[31] W. Chen *et al.*, "The Prognostic Value and Mechanisms of TMEM16A in Human Cancer," *Front Mol Biosci*, vol. 8, Feb. 2021, doi: 10.3389/fmolb.2021.542156.

[32] V. Cereda *et al.*, "New gene expressed in prostate: a potential target for T cell-mediated prostate cancer immunotherapy," *Cancer Immunology, Immunotherapy*, vol. 59, no. 1, pp. 63–71, Jan. 2010, doi: 10.1007/s00262-009-0723-6.

[33] M. Mohsenzadegan *et al.*, "Reduced expression of NGEP is associated with high-grade prostate cancers: a tissue microarray analysis," *Cancer Immunology, Immunotherapy*, vol. 62, no. 10, pp. 1609–1618, Oct. 2013, doi: 10.1007/s00262-013-1463-1.

[34] S. Vermeer *et al.*, "Targeted Next-Generation Sequencing of a 12.5 Mb Homozygous Region Reveals ANO10 Mutations in Patients with Autosomal-Recessive Cerebellar Ataxia," *The American Journal of Human Genetics*, vol. 87, no. 6, pp. 813–819, Dec. 2010, doi: 10.1016/j.ajhg.2010.10.015.

[35] M. Petkovic, J. Oses-Prieto, A. Burlinggame, L. Y. Jan, and Y. N. Jan, “TMEM16K is an interorganelle regulator of endosomal sorting,” 2019.

[36] L. Braga *et al.*, “Drugs that inhibit TMEM16 proteins block SARS-CoV-2 spike-induced syncytia,” *Nature*, 2021, doi: 10.1038/s41586-021-03491-6.

[37] P. J. Stansfeld *et al.*, “MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes,” *Structure*, vol. 23, no. 7, pp. 1350–1361, Jul. 2015, doi: 10.1016/j.str.2015.05.006.

[38] N. P. P. Bethel and M. Grabe, “Atomistic insight into lipid translocation by a TMEM16 scramblase,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 49, pp. 14049–14054, Dec. 2016, doi: 10.1073/pnas.1607574113.

[39] T. Jiang, K. Yu, H. C. Hartzell, and E. Tajkhorshid, “Lipids and ions traverse the membrane by the same physical pathway in the nhTMEM16 scramblase,” *Elife*, vol. 6, Sep. 2017, doi: 10.7554/eLife.28671.

[40] G. Khelashvili, X. Cheng, M. E. Falzone, M. Doktorova, A. Accardi, and H. Weinstein, “Membrane lipids are both the substrates and a mechanistically responsive environment of TMEM16 scramblase proteins,” *J Comput Chem*, 2020, doi: 10.1002/jcc.26105.

[41] G. Khelashvili, M. E. E. Falzone, X. Cheng, B.-C. C. Lee, A. Accardi, and H. Weinstein, “Dynamic modulation of the lipid translocation groove generates a conductive ion channel in Ca2+-bound nhTMEM16,” *Nat Commun*, vol. 10, no. 1, p. 4972, Dec. 2019, doi: 10.1038/s41467-019-12865-4.

[42] B.-C. C. Lee, G. Khelashvili, M. Falzone, A. K. K. Menon, H. Weinstein, and A. Accardi, “Gating mechanism of the extracellular entry to the lipid pathway in a TMEM16 scramblase,” *Nat Commun*, vol. 9, no. 1, p. 3251, Dec. 2018, doi: 10.1038/s41467-018-05724-1.

[43] X. Cheng, G. Khelashvili, and H. Weinstein, “The permeation of potassium ions through the lipid scrambling path of the membrane protein nhTMEM16,” *Front Mol Biosci*, vol. 9, Jul. 2022, doi: 10.3389/fmolb.2022.903972.

[44] A. Y. Kostritskii and J.-P. Machtens, “Molecular mechanisms of ion conduction and ion selectivity in TMEM16 lipid scramblases,” *Nat Commun*, vol. 12, no. 1, p. 2826, May 2021, doi: 10.1038/s41467-021-22724-w.

[45] M. Malvezzi *et al.*, “Out-of-the-groove transport of lipids by TMEM16 and GPCR scramblases,” *Proc Natl Acad Sci U S A*, 2018, doi: 10.1073/pnas.1806721115.

[46] C. J. Peters, H. Yu, J. Tien, Y. N. Jan, M. Li, and L. Y. Jan, “Four basic residues critical for the ion selectivity and pore blocker sensitivity of TMEM16A calcium-activated chloride channels,” *Proc Natl Acad Sci U S A*, 2015, doi: 10.1073/pnas.1502291112.

[47] C. J. Peters *et al.*, “The Sixth Transmembrane Segment Is a Major Gating Component of the TMEM16A Calcium-Activated Chloride Channel,” *Neuron*, 2018, doi: 10.1016/j.neuron.2018.01.048.

[48] C. Paulino *et al.*, “Structural basis for anion conduction in the calcium-activated chloride channel TMEM16A,” *Elife*, 2017, doi: 10.7554/eLife.26232.

[49] S. C. Le, Z. Jia, J. Chen, and H. Yang, “Molecular basis of PIP2-dependent regulation of the Ca2+-activated chloride channel TMEM16A,” *Nat Commun*, 2019, doi: 10.1038/s41467-019-11784-8.

[50] Z. Jia and J. Chen, “Specific PIP2 binding promotes calcium activation of TMEM16A chloride channels,” *Commun Biol*, 2021, doi: 10.1038/s42003-021-01782-2.

[51] A. K. M. Lam, S. Rutz, and R. Dutzler, “Inhibition mechanism of the chloride channel TMEM16A by the pore blocker 1PBC,” *Nat Commun*, 2022, doi: 10.1038/s41467-022-30479-1.

[52] K. Yu, T. Jiang, Y. Y. Cui, E. Tajkhorshid, and H. Criss Hartzell, “A network of phosphatidylinositol 4,5-bisphosphate binding sites regulates gating of the Ca2+-activated Cl− channel ANO1 (TMEM16A),” *Proc Natl Acad Sci U S A*, 2019, doi: 10.1073/pnas.1904012116.

[53] A. J. Lowry, P. Liang, Y. C. S. Wan, Z.-M. Pei, H. Yang, and Y. Zhang, “TMEM16 and TMEM63/OSCA proteins share a conserved potential to permeate ions and phospholipids.,” *bioRxiv*, Feb. 2024, doi: 10.1101/2024.02.04.578431.

[54] T. Le, Z. Jia, S. C. Le, Y. Zhang, J. Chen, and H. Yang, “An inner activation gate controls TMEM16F phospholipid scrambling,” *Nat Commun*, 2019, doi: 10.1038/s41467-019-09778-7.

[55] A. Kmit *et al.*, “Calcium-activated and apoptotic phospholipid scrambling induced by Ano6 can occur independently of Ano6 ion currents,” *Cell Death Dis*, vol. 4, no. 4, pp. e611–e611, Apr. 2013, doi: 10.1038/cddis.2013.135.

[56] T. W. Han *et al.*, “Chemically induced vesiculation as a platform for studying TMEM16F activity,” *Proceedings of the National Academy of Sciences*, vol. 116, no. 4, pp. 1309–1318, Jan. 2019, doi: 10.1073/pnas.1817498116.

[57] G. Khelashvili, E. Kots, X. Cheng, M. V. Levine, and H. Weinstein, “The allosteric mechanism leading to an open-groove lipid conductive state of the TMEM16F scramblase,” *Commun Biol*, vol. 5, no. 1, p. 990, Sep. 2022, doi: 10.1038/s42003-022-03930-8.

[58] K. Segawa, J. Suzuki, and S. Nagata, “Constitutive exposure of phosphatidylserine on viable cells,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 48, pp. 19246–19251, Nov. 2011, doi: 10.1073/pnas.1114799108.

[59] R. Watanabe, T. Sakuragi, H. Noji, and S. Nagata, “Single-molecule analysis of phospholipid scrambling by TMEM16F,” *Proc Natl Acad Sci U S A*, 2018, doi: 10.1073/pnas.1717956115.

[60] Z. Jia, J. Huang, and J. Chen, “Activation of TMEM16F by inner gate charged mutations and possible lipid/ion permeation mechanisms,” *Biophys J*, vol. 121, no. 18, pp. 3445–3457, Sep. 2022, doi: 10.1016/j.bpj.2022.08.011.

[61] M. E. Falzone *et al.*, “TMEM16 scramblases thin the membrane to enable lipid scrambling,” *Nat Commun*, vol. 13, no. 1, p. 2604, May 2022, doi: 10.1038/s41467-022-30300-z.

[62] M. E. Falzone *et al.*, “TMEM16 scramblases thin the membrane to enable lipid scrambling,” *Nat Commun*, vol. 13, no. 1, p. 2604, May 2022, doi: 10.1038/s41467-022-30300-z.

[63] C. Paulino, V. Kalienkova, A. K. M. Lam, Y. Neldner, and R. Dutzler, “Activation mechanism of the calcium-activated chloride channel TMEM16A revealed by cryo-EM,” *Nature*, 2017, doi: 10.1038/nature24652.

[64] M. Arndt, C. Alvadia, M. S. Straub, V. Clerico Mosina, C. Paulino, and R. Dutzler, “Structural basis for the activation of the lipid scramblase TMEM16F,” *Nat Commun*, vol. 13, no. 1, p. 6692, Nov. 2022, doi: 10.1038/s41467-022-34497-x.

[65] V. Kalienkova *et al.*, “Stepwise activation mechanism of the scramblase nhtmem16 revealed by cryo- em,” *Elife*, vol. 8, Feb. 2019, doi: 10.7554/eLife.44364.

[66] Z. Feng, O. E. Alvarenga, and A. Accardi, “Structural basis of closed groove scrambling by a TMEM16 protein,” *Nat Struct Mol Biol*, Apr. 2024, doi: 10.1038/s41594-024-01284-9.

[67] S. Dang *et al.*, “Cryo-EM structures of the TMEM16A calcium-activated chloride channel,” *Nature*, vol. 552, no. 7685, pp. 426–429, Dec. 2017, doi: 10.1038/nature25024.

[68] A. K. M. Lam, J. Rheinberger, C. Paulino, and R. Dutzler, “Gating the pore of the calcium-activated chloride channel TMEM16A,” *Nat Commun*, vol. 12, no. 1, p. 785, Feb. 2021, doi: 10.1038/s41467-020-20787-9.

[69] A. K. Lam and R. Dutzler, “Mechanistic basis of ligand efficacy in the calcium-activated chloride channel <scp>TMEM16A</scp>,” *EMBO J*, vol. 42, no. 24, Dec. 2023, doi: 10.15252/embj.2023115030.

[70] A. K. M. Lam, S. Rutz, and R. Dutzler, “Inhibition mechanism of the chloride channel TMEM16A by the pore blocker 1PBC,” *Nat Commun*, vol. 13, no. 1, p. 2798, May 2022, doi: 10.1038/s41467-022-30479-1.

[71] S. Feng *et al.*, “Cryo-EM Studies of TMEM16F Calcium-Activated Ion Channel Suggest Features Important for Lipid Scrambling,” *Cell Rep*, vol. 28, no. 2, pp. 567--579.e4, Jul. 2019, doi: 10.1016/j.celrep.2019.06.023.

[72] S. Feng *et al.*, “Identification of a drug binding pocket in TMEM16F calcium-activated ion channel and lipid scramblase,” *Nat Commun*, vol. 14, no. 1, p. 4874, Aug. 2023, doi: 10.1038/s41467-023-40410-x.

[73] V. Kalienkova, V. Clerico Mosina, and C. Paulino, “The Groovy TMEM16 Family: Molecular Mechanisms of Lipid Scrambling and Ion Conduction,” 2021. doi: 10.1016/j.jmb.2021.166941.

[74] P. Liang, Y. C. S. Wan, K. Yu, H. C. Hartzell, and H. Yang, “Niclosamide potentiates TMEM16A and induces vasoconstriction,” *Journal of General Physiology*, vol. 156, no. 7, Jul. 2024, doi: 10.1085/jgp.202313460.

[75] M. Tembo, K. L. Wozniak, R. E. Bainbridge, and A. E. Carlson, “Phosphatidylinositol 4,5-bisphosphate (PIP2) and Ca2+ are both required to open the Cl− channel TMEM16A,” *Journal of Biological Chemistry*, vol. 294, no. 33, pp. 12556–12564, Aug. 2019, doi: 10.1074/jbc.RA118.007128.

[76] A. Fiser and A. Sali, “ModLoop: Automated modeling of loops in protein structures,” *Bioinformatics*, 2003, doi: 10.1093/bioinformatics/btg362.

[77] J. Jumper *et al.*, “Highly accurate protein structure prediction with AlphaFold,” *Nature*, vol. 596, no. 7873, pp. 583–589, Aug. 2021, doi: 10.1038/s41586-021-03819-2.

[78] A. K. M. Lam and R. Dutzler, “Mechanism of pore opening in the calcium-activated chloride channel TMEM16A,” *Nat Commun*, vol. 12, no. 1, p. 786, Feb. 2021, doi: 10.1038/s41467-020-20788-8.

[79] A. K. M. Lam and R. Dutzler, “Mechanism of pore opening in the calcium-activated chloride channel TMEM16A,” *Nat Commun*, vol. 12, no. 1, p. 786, Feb. 2021, doi: 10.1038/s41467-020-20788-8.

[80] G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, and F. Noé, "Identification of slow molecular order parameters for Markov model construction," *J Chem Phys*, vol. 139, no. 1, Jul. 2013, doi: 10.1063/1.4811489.

[81] P. Deuflhard and M. Weber, "Robust Perron cluster analysis in conformation dynamics," *Linear Algebra Appl*, vol. 398, pp. 161–184, Mar. 2005, doi: 10.1016/j.laa.2004.10.026.

[82] M. P. Harrigan *et al.*, "MSMBuilder: Statistical Models for Biomolecular Dynamics," *Biophys J*, vol. 112, no. 1, pp. 10–15, Jan. 2017, doi: 10.1016/j.bpj.2016.10.042.

[83] J. M. Whitlock and H. C. Hartzell, "A Pore Idea: the ion conduction pathway of TMEM16/ANO proteins is composed partly of lipid," *Pflugers Arch*, vol. 468, no. 3, pp. 455–473, Mar. 2016, doi: 10.1007/s00424-015-1777-2.

[84] A. Šali and T. L. Blundell, "Comparative Protein Modelling by Satisfaction of Spatial Restraints," *J Mol Biol*, vol. 234, no. 3, pp. 779–815, Dec. 1993, doi: 10.1006/jmbi.1993.1626.

[85] M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, and J. H. Jensen, "PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K_a Predictions," *J Chem Theory Comput*, vol. 7, no. 2, pp. 525–537, Feb. 2011, doi: 10.1021/ct100578z.

[86] S. Jo, J. B. Lim, J. B. Klauda, and W. Im, "CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes," *Biophys J*, vol. 97, no. 1, pp. 50–58, Jul. 2009, doi: 10.1016/j.bpj.2009.04.013.

[87] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," *J Mol Graph*, 1996, doi: 10.1016/0263-7855(96)00018-5.

[88] M. J. Frisch *et al.*, "Gaussian~09 Revision E.01."

[89] M. J. Abraham *et al.*, "Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, 2015, doi: 10.1016/j.softx.2015.06.001.

[90] J. Huang and A. D. MacKerell, “CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data,” *J Comput Chem*, vol. 34, no. 25, pp. 2135–2145, Sep. 2013, doi: 10.1002/jcc.23354.

[91] J. Huang *et al.*, “CHARMM36m: an improved force field for folded and intrinsically disordered proteins,” *Nat Methods*, vol. 14, no. 1, pp. 71–73, Jan. 2017, doi: 10.1038/nmeth.4067.

[92] D. J. Evans and B. L. Holian, “The Nose–Hoover thermostat,” *J Chem Phys*, vol. 83, no. 8, pp. 4069–4074, Oct. 1985, doi: 10.1063/1.449071.

[93] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, “Molecular dynamics with coupling to an external bath,” *J Chem Phys*, vol. 81, no. 8, pp. 3684–3690, Oct. 1984, doi: 10.1063/1.448118.

[94] M. Parrinello and A. Rahman, “Polymorphic transitions in single crystals: A new molecular dynamics method,” *J Appl Phys*, vol. 52, no. 12, pp. 7182–7190, Dec. 1981, doi: 10.1063/1.328693.

[95] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, “LINCS: A linear constraint solver for molecular simulations,” *J Comput Chem*, vol. 18, no. 12, pp. 1463–1472, Sep. 1997, doi: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

[96] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, “A smooth particle mesh Ewald method,” *J Chem Phys*, vol. 103, no. 19, pp. 8577–8593, Nov. 1995, doi: 10.1063/1.470117.

[97] M. P. Harrigan *et al.*, “MSMBuilder: Statistical Models for Biomolecular Dynamics,” *Biophys J*, vol. 112, no. 1, pp. 10–15, Jan. 2017, doi: 10.1016/j.bpj.2016.10.042.

[98] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein, “MDAnalysis: A toolkit for the analysis of molecular dynamics simulations,” *J Comput Chem*, vol. 32, no. 10, pp. 2319–2327, Jul. 2011, doi: 10.1002/jcc.21787.

[99] P. Virtanen *et al.*, “SciPy 1.0: fundamental algorithms for scientific computing in Python,” *Nat Methods*, vol. 17, no. 3, pp. 261–272, Mar. 2020, doi: 10.1038/s41592-019-0686-2.

[100] O. S. Smart, J. G. Neduvelil, X. Wang, B. A. Wallace, and M. S. P. Sansom, “HOLE: A program for the analysis of the pore dimensions of ion channel structural models,” *J Mol Graph*, vol. 14, no. 6, pp. 354–360, 1996, doi: https://doi.org/10.1016/S0263-7855(97)00009-X.

[101] K. Yu, J. M. Whitlock, K. Lee, E. A. Ortlund, Y. Yuan Cui, and H. C. Hartzell, “Identification of a lipid scrambling domain in ANO6/TMEM16F,” *Elife*, vol. 4, Jun. 2015, doi: 10.7554/eLife.06901.

[102] T. Pomorski and A. K. Menon, “Lipid flippases and their biological functions,” *Cellular and Molecular Life Sciences*, vol. 63, no. 24, pp. 2908–2921, Dec. 2006, doi: 10.1007/s00018-006-6167-7.

[103] D. Li, C. Rocha-Roa, M. A. Schilling, K. M. Reinisch, and S. Vanni, “Lipid scrambling is a general feature of protein insertases,” *Proceedings of the National Academy of Sciences*, vol. 121, no. 17, Apr. 2024, doi: 10.1073/pnas.2319476121.

[104] Z. Feng, E. Di Zanni, O. Alvarenga, S. Chakraborty, N. Rychlik, and A. Accardi, “In or out of the groove? Mechanisms of lipid scrambling by TMEM16 proteins,” *Cell Calcium*, vol. 121, p. 102896, Jul. 2024, doi: 10.1016/j.ceca.2024.102896.

[105] F. R. Moss, J. Lincoff, M. Tucker, A. Mohammed, M. Grabe, and A. Frost, “Brominated lipid probes expose structural asymmetries in constricted membranes,” *Nat Struct Mol Biol*, vol. 30, no. 2, pp. 167–175, Feb. 2023, doi: 10.1038/s41594-022-00898-1.

[106] P. C. T. Souza *et al.*, “Martini 3: a general purpose force field for coarse-grained molecular dynamics,” *Nat Methods*, vol. 18, no. 4, pp. 382–388, Apr. 2021, doi: 10.1038/s41592-021-01098-3.

[107] S. Gyobu, K. Ishihara, J. Suzuki, K. Segawa, and S. Nagata, “Characterization of the scrambling domain of the TMEM16 family,” *Proceedings of the National Academy of Sciences*, vol. 114, no. 24, pp. 6274–6279, Jun. 2017, doi: 10.1073/pnas.1703391114.

[108] D. Li, C. Rocha-Roa, M. A. Schilling, K. M. Reinisch, and S. Vanni, “Lipid scrambling is a general feature of protein insertases,” *Proceedings of the National Academy of Sciences*, vol. 121, no. 17, Apr. 2024, doi: 10.1073/pnas.2319476121.

[109] Z. Feng, E. Di Zanni, O. Alvarenga, S. Chakraborty, N. Rychlik, and A. Accardi, “In or out of the groove? Mechanisms of lipid scrambling by TMEM16 proteins,” *Cell Calcium*, vol. 121, p. 102896, Jul. 2024, doi: 10.1016/j.ceca.2024.102896.

[110] W. Ye, T. W. Han, M. He, Y. N. Jan, and L. Y. Jan, “Dynamic change of electrostatic field in TMEM16F permeation pathway shifts its ion selectivity,” *Elife*, vol. 8, Jul. 2019, doi: 10.7554/eLife.45187.

[111] T. C. Anglin and J. C. Conboy, “Kinetics and Thermodynamics of Flip-Flop in Binary Phospholipid Membranes Measured by Sum-Frequency Vibrational Spectroscopy,” *Biochemistry*, vol. 48, no. 43, pp. 10220–10234, Nov. 2009, doi: 10.1021/bi901096j.

[112] W. Ye, T. W. Han, L. M. Nassar, M. Zubia, Y. N. Jan, and L. Y. Jan, “Phosphatidylinositol-(4, 5)-bisphosphate regulates calcium gating of small-conductance cation channel TMEM16F,” *Proc Natl Acad Sci U S A*, 2018, doi: 10.1073/pnas.1718728115.

[113] T. Tsuji *et al.*, “Predominant localization of phosphatidylserine at the cytoplasmic leaflet of the ER, and its TMEM16K-dependent redistribution,” *Proc Natl Acad Sci U S A*, 2019, doi: 10.1073/pnas.1822025116.

[114] Y. Han *et al.*, “Mechanical activation opens a lipid-lined pore in OSCA ion channels,” *Nature*, vol. 628, no. 8009, pp. 910–918, Apr. 2024, doi: 10.1038/s41586-024-07256-9.

[115] D. H. de Jong, S. Baoukina, H. I. Ingólfsson, and S. J. Marrink, “Martini straight: Boosting performance using a shorter cutoff and GPUs,” *Comput Phys Commun*, vol. 199, pp. 1–7, Feb. 2016, doi: 10.1016/j.cpc.2015.09.014.

[116] H. Kim, B. Fábián, and G. Hummer, “Neighbor List Artifacts in Molecular Dynamics Simulations,” *J Chem Theory Comput*, vol. 19, no. 23, pp. 8919–8929, Dec. 2023, doi: 10.1021/acs.jctc.3c00777.

[117] G. Bussi, D. Donadio, and M. Parrinello, “Canonical sampling through velocity rescaling,” *J Chem Phys*, vol. 126, no. 1, Jan. 2007, doi: 10.1063/1.2408420.

[118] E. Jurrus *et al.*, “Improvements to the <scp>APBS</scp> biomolecular solvation software suite,” *Protein Science*, vol. 27, no. 1, pp. 112–128, Jan. 2018, doi: 10.1002/pro.3280.

[119] T. Nugent and D. T. Jones, “Membrane protein orientation and refinement using a knowledge-based statistical potential,” *BMC Bioinformatics*, vol. 14, no. 1, p. 276, Dec. 2013, doi: 10.1186/1471-2105-14-276.

[120] P. C. Kroon *et al.*, “Martinize2 and Vermouth: Unified Framework for Topology Generation,” Nov. 2022.

[121] W. Kabsch and C. Sander, “Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features,” *Biopolymers*, vol. 22, no. 12, pp. 2577–2637, Dec. 1983, doi: 10.1002/bip.360221211.

[122] T. A. Wassenaar, H. I. Ingólfsson, R. A. Böckmann, D. P. Tieleman, and S. J. Marrink, “Computational Lipidomics with *insane* : A Versatile Tool for Generating Custom Membranes for Molecular Simulations,” *J Chem Theory Comput*, vol. 11, no. 5, pp. 2144–2155, May 2015, doi: 10.1021/acs.jctc.5b00209.

[123] R. Gowers *et al.*, “MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations,” 2016, pp. 98–105. doi: 10.25080/Majora-629e541a-00e.

[124] C. Szent-Gyorgyi *et al.*, “Malachite green mediates homodimerization of antibody VL domains to form a fluorescent ternary complex with singular symmetric interfaces,” *J Mol Biol*, vol. 425, no. 22, pp. 4595–4613, 2013, doi: 10.1016/j.jmb.2013.08.014.

[125] J. He *et al.*, "A genetically targetable near-infrared photosensitizer," *Nat Methods*, vol. 13, no. 3, pp. 263–268, 2016, doi: 10.1038/nmeth.3735.

[126] P. F. Aramendía, R. M. Negri, and E. S. Román, "(2 ') 2- 1]," pp. 3165–3173, 1994.

[127] E. M. S. Stennett, M. A. Ciuba, S. Lin, and M. Levitus, "Demystifying PIFE: The Photophysics behind the Protein-Induced Fluorescence Enhancement Phenomenon in Cy3," *Journal of Physical Chemistry Letters*, vol. 6, no. 10, pp. 1819–1823, 2015, doi: 10.1021/acs.jpclett.5b00613.

[128] J. Sturala *et al.*, "Excited-State Aromatic Interactions in the Aggregation-Induced Emission of Molecular Rotors," *J Am Chem Soc*, vol. 139, no. 49, pp. 17882–17889, 2017, doi: 10.1021/jacs.7b08570.

[129] Q. Li and Z. Li, "The Strong Light-Emission Materials in the Aggregated State: What Happens from a Single Molecule to the Collective Group," *Advanced Science*, vol. 4, no. 7, pp. 1–15, 2017, doi: 10.1002/advs.201600484.

[130] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, "Open Babel," *J Cheminform*, vol. 3, no. 33, pp. 1–14, 2011.

[131] Schrödinger LLC, "The PyMOL Molecular Graphics System, Version~1.8," Nov. 2015.

[132] J. Wang, W. Wang, P. A. Kollman, and D. A. Case, "Automatic atom type and bond type perception in molecular mechanical calculations," *J Mol Graph Model*, vol. 25, no. 2, pp. 247–260, 2006, doi: https://doi.org/10.1016/j.jmgm.2005.12.005.

[133] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general Amber force field," *J Comput Chem*, vol. 25, no. 9, pp. 1157–1174, 2004, doi: 10.1002/jcc.20035.

[134] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, "Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields," *J Phys Chem*, vol. 98, no. 45, pp. 11623–11627, Nov. 1994, doi: 10.1021/j100096a001.

[135] V. A. Rassolov, M. A. Ratner, J. A. Pople, P. C. Redfern, and L. A. Curtiss, "6-31G* basis set for third-row atoms," *J Comput Chem*, vol. 22, no. 9, pp. 976–984, Jul. 2001, doi: 10.1002/jcc.1058.

[136] F. Weigend, "Accurate Coulomb-fitting basis sets for H to Rn," *Physical Chemistry Chemical Physics*, vol. 8, no. 9, p. 1057, 2006, doi: 10.1039/b515623h.

[137] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems," *J Chem Phys*, vol. 98, no. 12, pp. 10089–10092, 1993.

[138] K.-H. Chow and D. M. Ferguson, "Isothermal-isobaric molecular dynamics simulations with Monte Carlo volume sampling," *Comput Phys Commun*, vol. 91, no. 1, pp. 283–289, 1995, doi: https://doi.org/10.1016/0010-4655(95)00059-O.

[139] J. Åqvist, P. Wennerström, M. Nervall, S. Bjelic, and B. O. Brandsdal, "Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm," *Chem Phys Lett*, vol. 384, no. 4, pp. 288–294, 2004, doi: https://doi.org/10.1016/j.cplett.2003.12.039.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
624C8C8F0A9E46F...
Author Signature

8/27/2024
Date