UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Structural and functional studies of putative ion channels

Permalink

https://escholarship.org/uc/item/5xj5m0f5

Author

Hoel, Christopher

Publication Date

2022

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Structural and functional studies of putative ion channels

By

Christopher Marston Hoel

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Stephen Brohawn, Chair
Professor Eva Nogales
Professor Diana Bautista
Professor Arash Komeili

Spring 2022

Abstract

Structural and functional studies of putative ion channels

by

Christopher Marston Hoel

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Stephen Brohawn, Chair

Ion channels are a large and diverse family of proteins broadly implicated in human physiology and that of other organisms. The study of the molecular mechanism of ion channels benefits from both "top-down" physiological and "bottom-up" biophysical studies. Here, we use a reductionist biophysical approach, with a focus on the use of cryo-electron microscopy (cryo-EM) to study the structure and function of three putative ion channels. The first, ORF3a is a putative ion channel encoded in the SARS-CoV-2 genome. Here, we determined the first ORF3a structure, and characterized ORF3a activity through electrophysiological recordings and fluorescent imaging of reconstituted ORF3a proteoliposomes. The second, Tweety-homolog family (TTYH) proteins, previously characterized as putative volume-regulated anion channels. Through parallel functional and structural studies, our work determined that TTYHs are not pore-forming subunits of anion channels. Additionally, we discovered a previously unknown "*trans*-dimeric" conformation of TTYH2. The third putative ion channel is TMEM87A, previously implicated in mechanosensitive conduction of cations. Here we determine the first structure of an TMEM87 protein, and through structural and bioinformatic analysis, identify related proteins forming the basis for a new protein family. Collectively, this work has demonstrated the value of cryo-EM for structural studies of small (<100 kDa) membrane proteins, while laying the foundation for deeper mechanistic understanding of each of these different proteins.

Acknowledgements

The opportunity to have the focus of one's time, energy, and effort be on the pursuit and advancement of knowledge is a privilege, and I am thankful to have had this opportunity during my time in graduate school. The MCB department here at Berkeley is a special place for many reasons, but especially for the prioritization of, and respect for, doing basic scientific research.

I am deeply grateful to have done my PhD with Steve Brohawn as my advisor. Steve has been a fantastic mentor -- he has deep scientific knowledge, an infectious passion for science, and is thoughtful and intentional about the environment in which we all do science. His support has been essential for my growth as a scientist, and I am so thankful of all he has done for me.

I also owe Tom Maresca tremendous thanks. Working in his lab at the University of Massachusetts Amherst prior to graduate school provided an invaluable scientific foundation that has enabled the work that I have done since. In addition to Tom's mentorship, I may not have found my way to science at all without the early support of Adrian Staub and Luke Remage-Healey at UMass, and appreciate both of their support.

Berkeley is a rich scientific community and I'm thankful to have received advice, friendship, and support from so many in my time here. I am extremely fortunate to have gone through graduate school in a cohort composed of many of the brightest, kindest, and hardest working people I have ever met. From commiseration to scientific troubleshooting, and plenty of fun throughout, this journey has been easier from knowing each of them. My thesis committee: Eva Nogales, Diana Bautista, and Arash Komeili have been invaluable. Our yearly meetings have provided important feedback, and I value each of their support and advice.

It has been great fun to join a new lab, help to shape the culture, and especially to watch how the lab has grown. The Brohawn lab is a fun, rigorous, and thrilling place to do science. I'm grateful to every member, past and present who have contributed to making the lab what it is.

Mom, Dad, Jonathan – I appreciate everything you've done for me. This opportunity and all that I have done here is thanks to your encouragement, love and support. Finally, for Victoria who has been there for all of it – thank you for everything and here's to the next adventure!

Table of contents

Chapter 1 Introduction to ion channels, the study of their structure and function, and cryo-electron microscopy

1.1 Ion channel diversity

Ion channels are ubiquitous across genomes, found in organisms from humans to bacteria, from plants to viruses.[1] Ion channels are macromolecular pores facilitating the permeation of ions across membranes. Critically, ion channels are ancient molecules, with significant diversity existing before the evolution of multicellular life, and continued expansion since (Fig. 1.1).[1,2,3] As a result, when examined as a class of proteins, ion channels display large functional diversity: channels can exhibit stunning selectivity for one ion over another, complex and multi-modal systems of regulation, and function in a wide array of physiological contexts.[4,5] In humans, ion channels play a key role in our health and when mutated are the cause of countless diseases.[6,7,8] As a result of their tremendous functional diversity, fundamental role in biology and physiology, and relevance to human health, an understanding of ion channels and their function is an important topic of study.

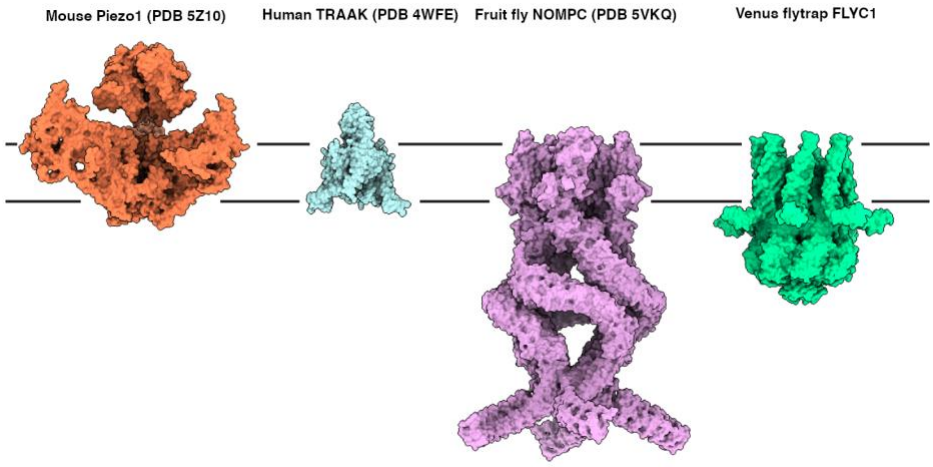


Figure 1.1: Diversity of ion channel structure. Surface representations of indicated mechanosensitive ion channel structures, representing a small subset of ion channel diversity. Mouse Piezo1 (PDB 5Z10) in orange, Human TRAAK (PDB 4WFE) in blue, Fruit fly NOMPC (PDB 5VKQ) in purple, and Venus flytrap FLYC1 (PDB 7N5E) in green.

The human genome contains hundreds of genes which code for the "pore-forming" subunit of ion channel complexes. This large number of diverse ion channel coding genes are necessary as a result of the wide range of physiological functions for which ion channels are involved.[3,6] Perhaps best studied with regard to their role in the action potential, ion channels are understood to be involved in every major aspect of our physiology, from renal function to the sensation of touch.[6,9] As a result, many functionally distinct ion channels are needed. For example, some ion channels allow potassium to pass but not sodium, while others do the opposite, some channels respond to changes in the cell's membrane potential while others respond to changes in pH. To accomplish these diverse physiological tasks, the corresponding ion channels have evolved distinct and complex mechanisms.[1,4,5,6]

We can appreciate this complexity through the recent discovery and subsequent characterization of the Piezo family of mechanosensitive ion channels.[10] Piezos were first discovered in 2010 through an effort to identify the molecular agent of a mechanically activated cation current found

in multiple cells.[11] In the decade following their initial identification, Piezos have been implicated in a wide range of physiological functions, including the sensation of touch, proprioception, force sensation in vasculature, and more.[10,12] Additionally, Piezo malfunction is implicated in a wide number of disorders, such as anemia and lymphatic dysplasia, and the knockout of either Piezo1 or Piezo2 in mice is lethal.[12,13] Recent cryo-EM structures of Piezo channels revealed a surprising and novel architecture for the protein, with the bulk of the complex formed by three large "blade" shaped domains (Fig. 1.1).[14,15,16] Intriguingly, while these structures have added to a great number of physiological studies and helped facilitate initial models for a mechanism of Piezo function, a full mechanistic understanding of Piezo function has been hindered by difficulty recording mechanosensitive currents from Piezo channels in a reductionist reconstituted system.[17,18]

In contrast with the case of Piezo – where the identification of the channels led to the rapid identification of their role in many physiological contexts, and despite our knowledge of the hundreds of ion channels encoded in the human genome, there are still multiple cases of well characterized currents for which the pore-forming molecule are not known. Perhaps most notably, in the hair cells of the inner ear a mechanosensitive current was identified in the 1970s, is known to be essential for the sensation of hearing, and despite decades of research the pore-forming subunit has yet to be conclusively identified, and a full understanding of the molecular mechanism remains to be revealed.[19,20,21] Notably, in multiple cases, channels have been proposed to be the pore-forming subunit of this ion channel complex, and later found not to be the case.[22] In the search for the identity of this channel, studies in native systems or of candidate genes expressed in heterologous cells have thus far been unable to conclusively identify the pore-forming subunit.[23,24] Here we can see the benefit of a reductionist and biophysical approach to study ion channel structure and function. Through the purification of candidate genes, determination of their structure, and functional measurements in a reconstituted system, unambiguous conclusions can be made regarding their function. In this case, perhaps such an approach will be necessary to finally answer this long-sought mystery.

1.2 Biophysical studies of ion channel structure and function

The benefit of a reductionist and biophysical approach to studying ion channel mechanism has most famously been demonstrated through the determination of the first ion channel crystal structure by Roderick MacKinnon's lab.[25] Not only was this discovery recognized with the award of the Nobel prize in Chemistry in 2003, but the determination of this structure allowed for a mechanistic understanding of potassium channel selectivity.[26] Before the determination of this first potassium channel structure, the field grappled with seemingly paradoxical observations: these potassium channels had been observed to conduct ions at rates nearing that of diffusion yet maintain exquisite selectivity – allowing potassium ions through while excluding sodium ions – how could this be the case?[25,26] The structure revealed that the potassium channel selectivity filter is arranged such that the backbone carbonyls line the pore of the channel and their positions mimic the arrangement of a hydrated potassium ion, allowing the rapid permeation of potassium, while excluding sodium due to its smaller hydrated radius.[25,26] Despite decades of research, this deep understanding of the protein's mechanism was inaccessible until the determination of the structure, demonstrating the importance of structural research for understanding ion channel mechanism.

Due to the tremendous diversity of ion channels and the complexity of their function our understanding of channel mechanisms remains surprisingly shallow despite decades of research and the clear importance of these proteins for human health, human physiology, and countless other organisms. This is certainly, in part, due to the historic challenge of determining the structures of membrane proteins, such as ion channels. However, recent technological advances in cryo-electron microscopy (cryo-EM) have dramatically accelerated the rate of protein structure determination, especially for membrane proteins (Fig. 1.2).[27]

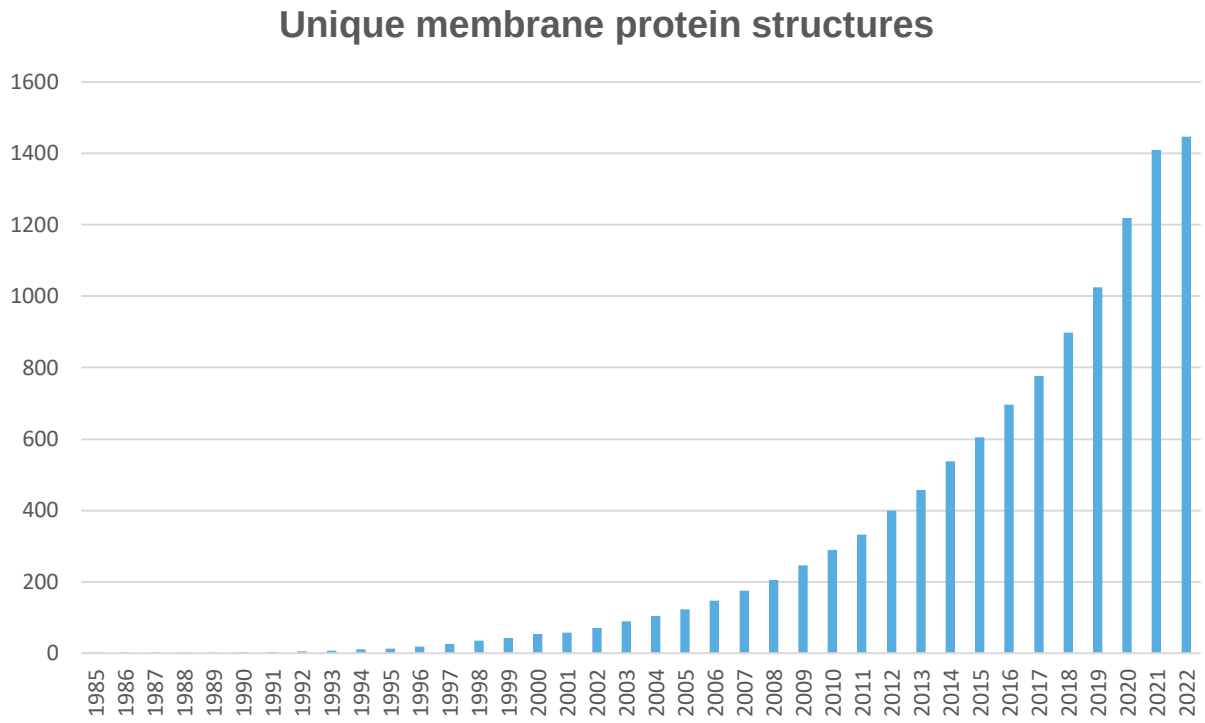


Figure 1.2: Unique membrane protein structures. Cumulative count of unique membrane protein structures deposited in the Protein Data Bank (PDB) with data from the Membrane Proteins of Known 3D Structure database (https://blanco.biomol.uci.edu/mpstruc/).

1.3 Cryo-EM's resolution revolution and membrane protein structure

The so called "resolution revolution" in cryo-EM has been transformative for structural biology, but especially so for structural studies of membrane proteins.[27,28] This advance was the cumulative result of multiple technological advances: in sample preparation, in microscope hardware, and in computational methods.[28] Together these improvements brought cryo-EM from an era sometimes called "blobology" (in reference to the extremely limited resolutions attainable) to the present time, when structures routinely achieve sufficient resolution for *de novo* modeling, and in some cases, have reached resolutions below 2 Å.[29,30] The use of cryo-EM alleviates many of the constraints which could make the determination of a membrane protein structure by crystallography intractable. Specifically, cryo-EM allows for sample preparation at much lower concentrations, does not require the growth of crystals, and allows for greater sample heterogeneity.[31]

The use of cryo-EM to study membrane protein structure has a rich history.[32,33] Beginning well before the resolution revolution with Henderson and Unwin's sub-nanometer map of bacteriorhodopsin in 1981, the first protein complex structure determined by electron microscopy.[32,33] Additionally, of the small handful of structures which are synonymous with the beginning of the resolution revolution, one of the most notable is the cryo-EM structure of the transient receptor potential (TRP) subfamily V member 1 ion channel by the Julius and Cheng labs in 2013 to 3.4 Å.[34] The first TRP currents were recorded in *Drosophilia* photoreceptor cells in 1977 and it took until the resolution revolution of cryo-EM to determine the first TRP channel structure almost 40 years later.[34,35,36] In the 8 years following this initial cryo-EM structure, over 100 TRP channel structures have been determined by cryo-EM.[37] These structures have not only featured additional members of the family, but have also examined small molecule binding, the effects of different temperatures, different pH, and of diseasing causing mutations.[37] This incredible bounty of structural information has allowed a deeper and more mechanistic understanding of TRP channel function. The benefit of cryo-EM is, of course, not limited to the study of TRP channels, and the technique has facilitated the determination of countless other membrane protein structures (Fig. 1.3).[33]

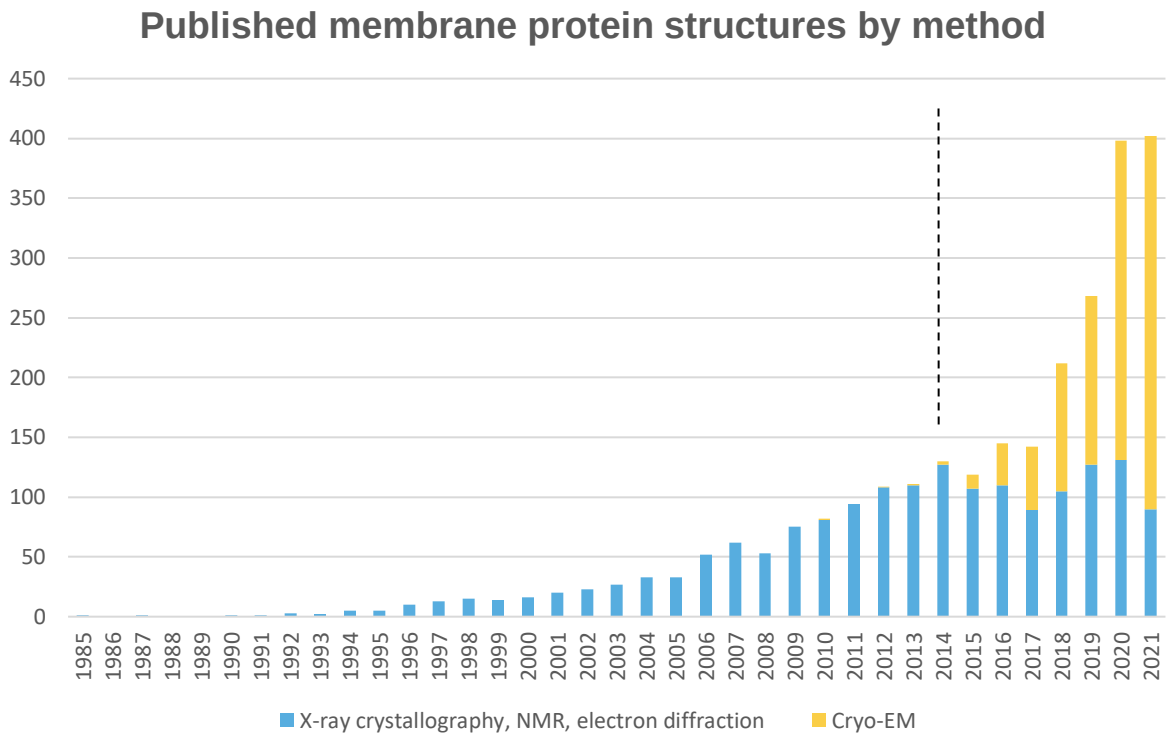


Figure 1.3: Published membrane protein structures by method. Stacked counts of membrane protein structures deposited in the Protein Data Bank (PDB), plotted using data from the Membrane Proteins of Known 3D Structure database (https://blanco.biomol.uci.edu/mpstruc/). Blue bars represent structures determined with methods other than cryo-EM (X-ray crystallography, nuclear magnetic resonance (NMR), and electron diffraction), orange bars represent structures determined by cryo-EM. The dashed black line corresponds to the period termed the "resolution revolution" in cryo-EM.

1.4 Lipid nanodiscs and their benefit

The development and use of protein lipid nanodiscs has significantly benefitted the determination of membrane protein structures by cryo-EM, and is especially useful for small membrane proteins.[38,39] Lipid nanodiscs have been developed by the Sligar lab and use engineered apolipoproteins termed membrane scaffold proteins (MSP) to create puck like lipid bilayers contained within an MSP of desired radius.[40] Intuitively, the bilayer of the lipid nanodisc provides an environment which is more physiologically relevant than that of a detergent micelle. This is especially relevant for proteins which are regulated by lipids or have specific protein-lipid interactions. Experimentally, lipid nanodiscs have been shown to improve the thermodynamic stability of purified proteins, and provide additional control over experimental conditions as the lipid composition of the nanodisc can be manipulated as desired.[38,39,40] Following initial studies which established the use of nanodiscs in modern cryo-EM, the use of nanodiscs has become more routine for membrane protein structure determination by cryo-EM.[38,39,41] Intriguingly, our group found the use of nanodiscs critical to resolve the structure of mouse KCC4 (~120 kDa protein, with resolved mass corresponding to ~55 kDa) by cryo-EM.[42] In addition to the potential improvement to protein stability in nanodiscs, we hypothesize that nanodiscs provide better signal-to-noise than detergent micelles. We find that for structure determination of small membrane proteins (< 100 kDa) "contaminating" empty detergent micelles can make early classification of particle images intractable, and is alleviated by the use of nanodiscs.[43] Nanodiscs are not the only tool developed to aid in the structure determination of small membrane proteins, other important tools have been the use of protein fiducial markers such as nanobodies, antibody fragment antigen-binding regions (Fab), and related proteins.[43]

1.5 Structural and functional studies of small putative ion channels by cryo-EM

Driven by the innovations in cryo-EM and the opportunity to leverage reductionist methods to understand ion channel mechanism, in the present work I investigate three putative ion channels. Each of these proteins had been minimally characterized and are well poised to yield mechanistic insight through the application of biophysical methods such as cryo-EM and reconstitution-based functional assays. Chapter 2 describes studies performed in response to the coronavirus disease 2019 (COVID-19) pandemic, focused on the ORF3a gene encoded in the SARS-CoV-2 genome, as prior work suggested that the related ORF3a of SARS-CoV-1 functioned as a cation channel. Chapter 3 describes structural and functional studies of tweety-homolog (TTYH) family proteins, long believed to function as calcium-activated or volume-regulated anion channels. Chapter 4 describes work done to study TMEM87A, described in one study as a member of a novel mechanosensitive transduction pathway and both sufficient and necessary for the observation of this novel mechanosensitive current. Chapter 5 provides concluding remarks regarding these studies collectively as a body of work, demonstrating the value of cryo-EM for structural and functional studies of poorly characterized putative ion channels.

1.6 References

1. Anderson, P. A. V. & Greenberg, R. M. Phylogeny of ion channels: clues to structure and function. *Comparative Biochemistry and Physiology* **129,** 17-18 (2001).

2. Brehm, P., Okamura, Y. & Mandel, G. Ion channel evolution. *Seminars in Neuroscience* **3**, 355–367 (1991).
3. Liebeskind, B. J., Hillis, D. M. & Zakon, H. H. Convergence of ion channel genome content in early animal evolution. *Proc. Natl. Acad. Sci. U.S.A.* **112**, (2015).
4. Roux, B. Ion channels and ion selectivity. *Essays in Biochemistry* **61**, 201–209 (2017).
5. Ramsey, I. S., Delling, M. & Clapham, D. E. An introduction to TRP channels. *Annu. Rev. Physiol.* **68**, 619–647 (2006).
6. Ackerman, M. J. & Clapham, D. E. Ion Channels — Basic Science and Clinical Disease. *The New England Journal of Medicine* 12 (1997).
7. Cooper, E. C. & Jan, L. Y. Ion channel genes and human neurological disease: Recent progress, prospects, and challenges. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4759–4766 (1999).
8. Amin, A. S., Tan, H. L. & Wilde, A. A. M. Cardiac ion channels in health and disease. *Heart Rhythm* **7**, 117–126 (2010).
9. Schwiening, C. J. A brief historical perspective: Hodgkin and Huxley: Classical Perspectives. *The Journal of Physiology* **590**, 2571–2575 (2012).
10. Wu, J., Lewis, A. H. & Grandl, J. Touch, Tension, and Transduction – The Function and Regulation of Piezo Ion Channels. *Trends in Biochemical Sciences* **42**, 57–71 (2017).
11. Coste, B. *et al.* Piezo proteins are pore-forming subunits of mechanically activated channels. *Nature* **483**, 176–181 (2012).
12. Murthy, S. E., Dubin, A. E. & Patapoutian, A. Piezos thrive under pressure: mechanically activated ion channels in health and disease. *Nat Rev Mol Cell Biol* **18**, 771–783 (2017).
13. Beech, D. J. & Kalli, A. C. Force Sensing by Piezo Channels in Cardiovascular Health and Disease. *ATVB* **39**, 2228–2239 (2019).
14. Guo, Y. R. & MacKinnon, R. Structure-based membrane dome mechanism for Piezo mechanosensitivity. *eLife* **6**, e33660 (2017).
15. Saotome, K. *et al.* Structure of the mechanically activated ion channel Piezo1. *Nature* **554**, 481–486 (2018).
16. Yang, X. *et al.* Structure deformation and curvature sensing of PIEZO1 in lipid membranes. *Nature* **604**, 377–383 (2022).
17. Cox, C. D. *et al.* Removal of the mechanoprotective influence of the cytoskeleton reveals PIEZO1 is gated by bilayer tension. *Nat Commun* **7**, 10366 (2016).
18. Syeda, R. *et al.* Piezo1 Channels Are Inherently Mechanosensitive. *Cell Reports* **17**, 1739–1746 (2016).
19. Corey, D. P. & Hudspeth, A. J. Ionic basis of the receptor potential in a vertebrate hair cell. *Nature* **281**, 675–677 (1979).
20. Fettiplace, R. Hair Cell Transduction, Tuning, and Synaptic Transmission in the Mammalian Cochlea. in *Comprehensive Physiology* (ed. Terjung, R.) 1197–1227 (Wiley, 2017). doi:10.1002/cphy.c160049.
21. Cunningham, C. L. & Müller, U. Molecular Structure of the Hair Cell Mechanoelectrical Transduction Complex. *Cold Spring Harb Perspect Med* **9**, a033167 (2019).
22. Corey, D. P. What is the hair cell transduction channel?: The hair cell transduction channel. *The Journal of Physiology* **576**, 23–28 (2006).
23. Cunningham, C. L. *et al.* TMIE Defines Pore and Gating Properties of the Mechanotransduction Channel of Mammalian Cochlear Hair Cells. *Neuron* **107**, 126-143.e8 (2020).

24. Zheng, W. & Holt, J. R. The Mechanosensory Transduction Machinery in Inner Ear Hair Cells. *Annu. Rev. Biophys.* **50**, 31–51 (2021).
25. Doyle, D. A. *et al.* The Structure of the Potassium Channel: Molecular Basis of K^+ Conduction and Selectivity. *Science* **280**, 69–77 (1998).
26. MacKinnon, R. Potassium Channels and the Atomic Basis of Selective Ion Conduction (Nobel Lecture). *Angew. Chem. Int. Ed.* **43**, 4265–4277 (2004).
27. Cheng, Y. Membrane protein structural biology in the era of single particle cryo-EM. *Current Opinion in Structural Biology* **52**, 58–63 (2018).
28. Kühlbrandt, W. The Resolution Revolution. *Science* **343**, 1443–1444 (2014).
29. Nakane, T. *et al.* Single-particle cryo-EM at atomic resolution. *Nature* **587**, 152–156 (2020).
30. Yip, K. M., Fischer, N., Paknia, E., Chari, A. & Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* **587**, 157–161 (2020).
31. Cheng, Y. Single-Particle Cryo-EM at Crystallographic Resolution. *Cell* **161**, 450–457 (2015).
32. Cheng, Y. Single-particle cryo-EM—How did it get here and where will it go. *Science* **361**, 876–880 (2018).
33. Kühlbrandt, W. Forty years in cryoEM of membrane proteins. *Microscopy* **71**, i30–i50 (2022).
34. Liao, M., Cao, E., Julius, D. & Cheng, Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **504**, 107–112 (2013).
35. Minke, B. Drosophila mutant with a transducer defect. *Biophys. Struct. Mechanism* **3**, 59–64 (1977).
36. Minke, B. The History of the *Drosophila* TRP Channel: The Birth of a New Channel Superfamily. *Journal of Neurogenetics* **24**, 216–233 (2010).
37. Zhao, Y., McVeigh, B. M. & Moiseenkova-Bell, V. Y. Structural Pharmacology of TRP Channels. *Journal of Molecular Biology* **433**, 166914 (2021).
38. Denisov, I. G. & Sligar, S. G. Nanodiscs for structural and functional studies of membrane proteins. *Nat Struct Mol Biol* **23**, 481–486 (2016).
39. Efremov, R. G., Gatsogiannis, C. & Raunser, S. Lipid Nanodiscs as a Tool for High-Resolution Structure Determination of Membrane Proteins by Single-Particle Cryo-EM. in *Methods in Enzymology* vol. 594 1–30 (Elsevier, 2017).
40. Ritchie, T. K. *et al.* Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs. in *Methods in Enzymology* vol. 464 211–231 (Elsevier, 2009).
41. Gao, Y., Cao, E., Julius, D. & Cheng, Y. TRPV1 structures in nanodiscs reveal mechanisms of ligand and lipid action. *Nature* **534**, 347–351 (2016).
42. Reid, M. S., Kern, D. M. & Brohawn, S. G. Cryo-EM structure of the potassium-chloride cotransporter KCC4 in lipid nanodiscs. *eLife* **9**, e52505 (2020).
43. Nygaard, R., Kim, J. & Mancia, F. Cryo-electron microscopy analysis of small membrane proteins. *Current Opinion in Structural Biology* **64**, 26–33 (2020).

Chapter 2 Cryo-EM structure of SARS-CoV-2 ORF3a in lipid nanodiscs

This chapter is adapted, with permission, from work previously published: Kern, D.M., Sorum, B., Mali, S.S., Hoel, C.M., Sridharan, S., Remis, J.P., Toso, D.B., Kotecha, A., Bautista, D.M., Brohawn, S.G. Cryo-EM structure of SARS-CoV-2 ORF3a in lipid nanodiscs. *Nat Struct Mol Biol* **28,** 573–582 (2021). https://doi.org/10.1038/s41594-021-00619-0

2.1 Abstract

SARS-CoV-2 ORF3a is a putative viral ion channel implicated in autophagy inhibition, inflammasome activation and apoptosis. 3a protein and anti-3a antibodies are found in infected patient tissues and plasma. Deletion of 3a in SARS-CoV-1 reduces viral titer and morbidity in mice, suggesting it could be an effective target for vaccines or therapeutics. Here, we present structures of SARS-CoV-2 3a determined by cryo-EM to 2.1-Å resolution. 3a adopts a new fold with a polar cavity that opens to the cytosol and membrane through separate water- and lipid-filled openings. Hydrophilic grooves along outer helices could form ion-conduction paths. Using electrophysiology and fluorescent ion imaging of 3a-reconstituted liposomes, we observe Ca2+-permeable, nonselective cation channel activity, identify mutations that alter ion permeability and discover polycationic inhibitors of 3a activity. 3a-like proteins are found across coronavirus lineages that infect bats and humans, suggesting that 3a-targeted approaches could treat COVID-19 and other coronavirus diseases.

2.2 Introduction

Coronavirus disease 2019 (COVID-19), caused by the SARS-CoV-2 virus, is an ongoing global pandemic. Vaccine and therapeutic development are predominantly focused on the essential virus-encoded Spike, main protease and RNA-dependent RNA polymerase proteins. High-resolution structures of these targets, some in complex with drug candidates or neutralizing antibodies, have yielded mechanistic insight into their function and have provided a platform for successful structure-guided vaccine and drug design[1,2,3,4,5,6]. Still, expanding our knowledge of SARS-CoV-2 target proteins remains important both for understanding SARS-CoV-2 virology and developing alternative treatments to mitigate against potential resistance that evolves, or new viruses that emerge, in the future[7].

Virally encoded open reading frame 3a (ORF3a, 3a for short) has been considered a potential target for vaccines or therapeutics[8,9]. 3a is highly conserved within the *Betacoronavirus* subgenus *Sarbecovirus*, which includes SARS-CoV-1 and other related bat coronaviruses (Supplemental Figure 2.1)[10]. Biopsies from SARS-CoV-1 patients show 3a expression in infected tissue and SARS-CoV-1 and SARS-CoV-2 patient plasma contain anti-3a antibodies[8,11,12,13]. SARS-CoV-1 3a has been implicated in inflammasome activation[14] and both apoptotic[15] and necrotic cell death[16] while SARS-CoV-2 3a has been implicated in apoptosis and inhibition of autophagy in vitro[17]. In mouse models of both SARS-CoV-1 and 2 infection, genomic deletion of 3a reduces viral titer and morbidity[18,19].

3a has been proposed to form an ion channel and is one of two putative viroporins encoded by the SARS-CoV-2 genome together with the envelope protein E (SARS-CoV-1 encodes an additional putative viroporin, ORF8a). SARS-CoV-1 3a has been reported to form an emodin- and Ba^{2+}-sensitive cation channel[20,21]. Viroporins from other viruses have been shown to modify host membrane permeability to promote viral assembly and release[22,23,24]. However, the precise roles of 3a in disease pathogenesis are unknown, precluded in part by the lack of a mechanistic understanding of 3a. Here, we determine structures of SARS-CoV-2 3a by cryo-electron microscopy (cryo-EM) and assess putative ion channel activity in vitro using electrophysiology and fluorescent ion flux assays.

2.3 Results

2.3.1 Structure determination

We determined structures of dimeric and tetrameric SARS-CoV-2 3a in lipid nanodiscs. Full length SARS-CoV-2 3a was heterologously expressed in *Spodoptera frugiperda* (Sf9) cells with a cleavable C-terminal green fluorescent protein (GFP) tag. Purification of 3a in detergent resulted in two species separable by gel filtration (Supplemental Figure 2.2). Most 3a runs at a position consistent with a 62-kDa dimer (Supplemental Figure 2.2a,c,e) and roughly 10–30% runs as a 124-kDa tetramer (Supplemental Figure 2.2). A similar degree of tetramer formation was observed at low concentrations of 3a by fluorescence size-exclusion chromatography, indicative of a biochemically stable species rather than concentration-dependent nonspecific aggregation (Supplemental Figure 2.2e). These data are consistent with previous reports of dimeric and tetrameric SARS-CoV-1 3a observed by western blot[16,20].

We separately reconstituted dimeric and tetrameric SARS-CoV-2 3a into nanodiscs made from the scaffold protein MSP1E3D1 and a mixture of DOPE, POPC and POPS lipids, and determined their structures by cryo-EM to 2.9- and 6.5-Å resolution (Supplemental Figures 2.3-2.6 and Table 2.1). Based on a report of improved resolution of apoferritin and $GABA_A$ structures using new cryo-EM instrumentation (including the electron source, energy filter and camera; see Methods for details)[25], we asked whether similar improvements would be possible for the sub-100-kDa 3a. Indeed, a reconstruction generated from data collected on this instrumentation, using the same batch of grids, was improved to 2.1 Å (Fig. 2.1, Supplemental Figure 2.3, 2.8, 2.9 and Table 2.1). We note we cannot exclude other differences including in accelerating voltage, pixel size or defocus range from potentially contributing to the improvement in resolution.

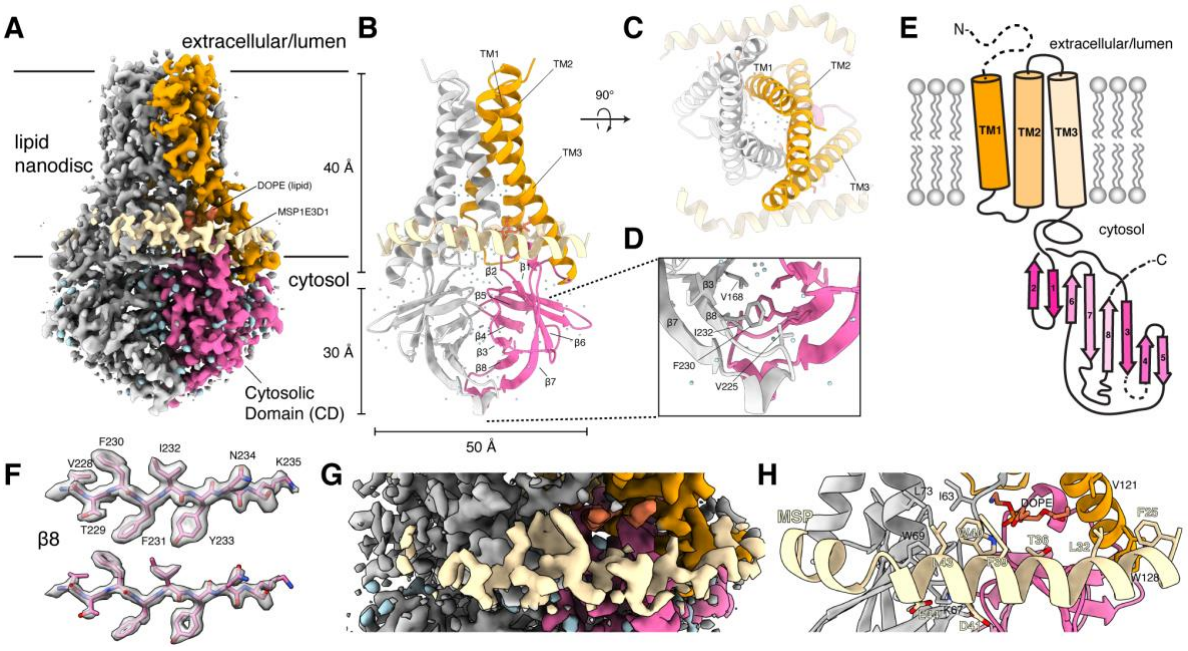


Figure 2.1: Structure of dimeric 3a in lipid nanodiscs. a, Cryo-EM map of the 3a dimer in MSP1E3D1 nanodiscs at 2.1-Å nominal resolution, viewed from the membrane plane. One subunit is colored gray, and the second subunit is colored with transmembrane region in orange and the CD in pink. Density from the nanodisc MSP1E3D1 is colored tan, and the DOPE lipid is coral. **b**,**c**, Model of dimeric 3a viewed from the membrane (**b**) (as in **a**) and from

the extracellular or lumenal side (**c**). **d**, Enlarged view of the interaction between subunits in the CD with residues forming the hydrophobic core indicated. **e**, Schematic of a 3a monomer. Secondary structure elements are indicated, and unmodeled termini and a five-amino-acid β3–β4 loop are shown with dashed lines. **f**, Cryo-EM density and model for a selected strand from the CD at two different thresholds. **g**, Enlarged view of density in the MSP1E3D1-interaction region. **h**, Model in the same region as **g** with key residues in the area displayed as sticks.

2.3.2 3a dimer structure

The 2.9-Å dimeric reconstruction (with C2 symmetry applied, PDB 6XDC) permitted de novo modeling of 195 of the 275 amino acids per protomer chain. The N terminus (amino acids 1–39), C terminus (amino acids 239–275) and a short cytoplasmic loop (amino acids 175–180) are weakly resolved in the density map due to conformational differences between particles or because they are disordered (Supplemental Figure 2.4 and 2.5). The 2.1-Å dimeric reconstruction (with C2 symmetry applied, PDB 7KJR) allowed for improved rotamer assignments due to better defined side chain density throughout the protein, including visible holes in many aromatic residues (Fig. 2.1a-f). We additionally modeled a portion of each MSP1E3D1 scaffold protein (amino acids 25–55, Fig. 2.1g,h), two DOPE lipids and 122 water molecules. We refer to the higher-resolution structure in the remainder of the paper unless otherwise indicated. The scaffold protein is partially well-resolved due to a specific interaction visible on each side of the 3a dimer (even in C1 reconstructions) between MSP1E3D1 W40 and the transmembrane region of 3a (Fig. 2.1g,h). The symmetric nature of this interaction means that the two MSPE3D1 protomers must twist around the lipid bilayer rather than adopting the canonical arrangement of two parallel stacked rings.

3a adopts a new fold that, to our knowledge, has not been observed in available protein structures. Querying the protein structure database for homologs with Dali returned only weak hits for fragments of 3a domains[26]. Viewed from the membrane plane, 3a is approximately 70 Å tall with a 40-Å high transmembrane region and a cytosolic domain (CD) extending 30 Å out of the membrane (Fig. 2.1a,b). The transmembrane region is composed of three helices per protomer with N termini oriented on the extracellular or lumenal side and C termini on the cytosolic side of the membrane. Viewed from the extracellular side, the transmembrane helices (TMs) trace the circumference of an ellipse with TM1–3 from one protomer followed by TM1–3 of the second protomer in a clockwise order (Fig. 2.1c). TM1s and TM2s pack against each other across the elliptical minor axis with TM3s positioned at the main axis vertices. TM1–TM2 and TM2–TM3 are joined by short intracellular and extracellular linkers, respectively.

The transmembrane region connects to the CD through a turn-helix-turn motif following TM3. Each protomer chain forms a pair of opposing β-sheets packed against one another in an eight stranded β-sandwich (Fig. 2.1b,d). The outer sheet is formed by strands β1, β2 and β6 and the N-terminal half of β7. The inner sheet is formed by strands β3, β4, β5 and β8 and the C-terminal half of β7. The inner sheets from each protomer interact through a large (roughly 940 $Å^2$ of buried surface area per chain) and highly complementary interface with residues V168, V225, F230 and I232 forming a continuous hydrophobic core (Fig. 2.1d). The interaction between β-sandwiches from each protomer thus forms a strong and stable link between monomers in the dimer.

2.3.3 Structural features of 3a tetramers

We next examined the structure of 3a tetramers. Two-dimensional (2D) class averages of tetrameric 3a show a side-by-side arrangement of two dimers with separated TMs and close juxtaposition of CDs (Fig. 2.2a). Tetrameric 3a reconstructions had lower final resolutions (roughly 6.5 Å) than dimeric 3a (Fig. 2.2a and Supplemental Figure 2.3), which precluded detailed model building, but were sufficiently featured in the CDs for rigid-body docking of two copies of the 3a dimer to determine their approximate relative orientation (Fig. 2.2b and Supplemental Figure 2.10a). The best fit model shows a continuous interface (roughly 300 $Å^2$ buried surface area per dimer) formed between TM3–CTD linkers and β1–β2 linkers from neighboring dimers. While the exact positions of side chains cannot be determined at this resolution, residues W131, R134, K136, H150, T151, N152, C153 and D155 are poised to mediate tetramerization through a network of hydrophobic, polar and electrostatic interactions (Fig. 2.2c). In SARS-CoV-1 3a, reducing agents and a C133A mutation resulted in loss of oligomerization, membrane localization and ion channel activity[20]. However, expression of the C133A mutant was dramatically reduced, suggesting these results might be a consequence of protein destabilization[20]. In SARS-CoV-2 3a, C133 is located in a notable cysteine-rich pocket adjacent to the tetramerization interface (Supplemental Figure 2.10b). At the base of TM3, C133 projects back toward the top faces of β1 and β2 in close proximity to solvent exposed C148 and buried C157. All three cysteines are reduced in the structure, although they are within potential disulfide-bonding distance (Cα distance 4.4–6.5 Å) (Supplemental Figure 2.10b). While it is unlikely a disulfide involving C133 mediates tetramerization of 3a without substantial rearrangement of this region, it could be that disruption of this cysteine-rich pocket with cysteine modifying agents or mutations disfavors 3a oligomerization allosterically.

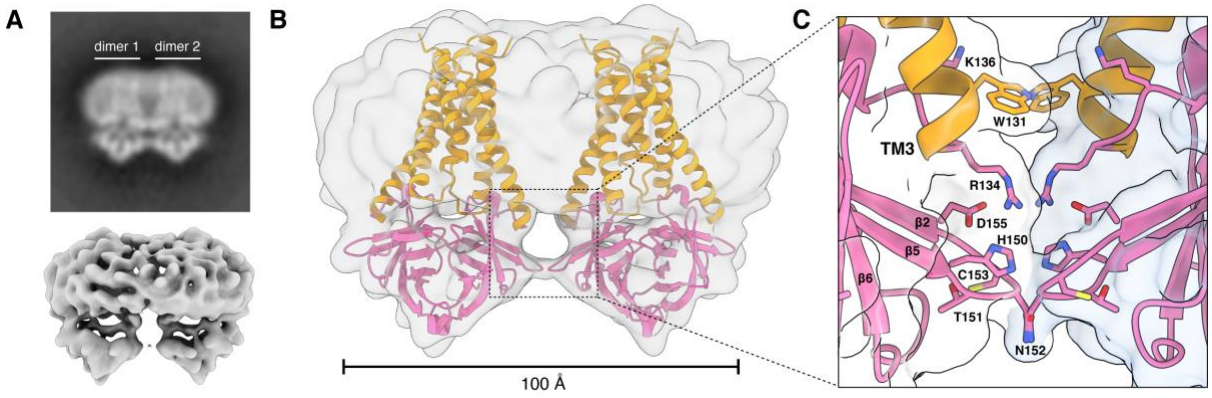


Figure 2.2: Structure of tetrameric 3a in lipid nanodiscs. a, Two-dimensional class average of tetrameric 3a in MSP1E3D1 lipid nanodiscs (top) and cryo-EM map at 6.5-Å nominal resolution (bottom). **b**, Approximate assembly of 3a tetramers. Two 3a dimer structures are rigid-body fit into the tetrameric cryo-EM map. **c**, Enlarged view of the interface between 3a dimers observed after rigid-body fitting with residues potentially mediating the interaction displayed as sticks.

2.3.4 Potential ion permeation paths through 3a

Ion conduction pathways are required for ion channel function, so we analyzed the 3a structures to identify potential permeation routes. 3a forms a large polar cavity within the inner half of the TM region. The cavity is continuous with the cytosol and surrounding bilayer through three pairs of openings: the upper, lower and intersubunit tunnels (Fig. 2.3a-f). The upper tunnels are

formed between TM2 and TM3 within each protomer, narrow to roughly 2.2 Å in radius, and likely open to the membrane (Fig. 2.3a-c). The lower tunnels run underneath the TM1–TM2 linker and above the CD, narrow to roughly 2 Å in radius, and open to the cytosol (Fig. 2.3a,d,e). The lower tunnels are open paths for ion and water movement between the cell interior and channel cavity. Consistently, a network of ordered water molecules extends through these openings, the cavity and the CD (Fig. 2.3b,d). The intersubunit tunnels run between TM1 and TM3 from opposing protomers, just above the CD, and open to the membrane–cytosol interface (Fig. 2.3a,f). Strong lipid-shaped density present within each intersubunit tunnel is modeled as DOPE (the most abundant lipid in these nanodiscs) with the ethanolamine head group pointed into the cavity. Lipid binding is stabilized by electrostatic interactions: the positively charged ethanolamine interacts with negatively charged D142 and the backbone carbonyls of I63 and D142 while the lipid phosphate interacts with R126, R122, Y206, N144 and the backbone amide of L65.

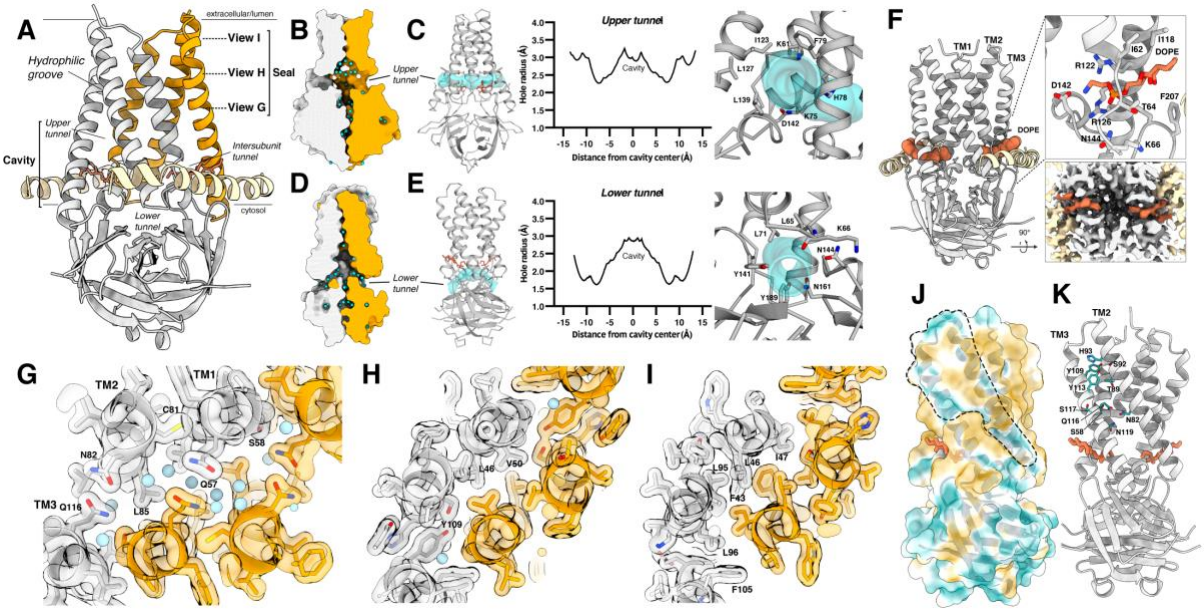


Figure 2.3: The 3a polar cavity and tunnels. a, View of a 3a dimer from the membrane plane with key regions labeled and with planes (dotted lines) and viewing direction (arrow) for **g**–**i**) indicated. **b**, The 3a upper tunnel with solvent-excluded surface shown for each subunit and water molecules shown as light blue spheres. The approximate lipid bilayer region is marked by black lines. One 3a subunit is colored orange and the other in gray. **c**, Left, model of 3a (gray) with the HOLE path through the upper tunnel shown in transparent cyan. Lipid (coral) is displayed as sticks. Middle, the radius of the upper tunnel as a function of distance from the cavity center calculated with HOLE. Right, the view into the upper tunnel from the membrane with key residues shown as sticks. **d**, As in **b**, for a view of the lower tunnel. **e**, As in **c**, for the lower tunnel, with the view on the right into the tunnel from the cytosol. **f**, The intersubunit tunnel with bound DOPE lipid (coral) shown as space-filling spheres. Top right, an enlarged view of the lipid interaction with key residues shown as sticks. Bottom-right, a cut-through view of the cryo-EM density of the lipid interaction region viewed from the cytosolic side. **g**–**i**, Model and transparent surface of the viewing planes indicated in **a** (**g**, view **G**; **h**, view **H**; **i**, view **I**). Water molecules are shown as light blue spheres. One 3a subunit is colored orange and the other in gray. **j**, The solvent-excluded surface of 3a colored from hydrophilic (dark cyan) to most hydrophobic (dark orange). A membrane-facing hydrophilic groove is outlined. **k**, Key residues in the hydrophilic groove are depicted as sticks colored in dark cyan.

While the polar cavity reduces the energetic barrier to ion movement across the inner half of the low dielectric membrane, the current structures lack a central pore connecting the cavity to the lumenal or extracellular side. Instead, a hydrophobic seal is formed between TMs in the extracellular half of the membrane above the polar cavity (Fig. 2.3). The top of the cavity is formed by polar interactions between Q57, S58, N82 and Q116 side chains (Fig. 2.3a,g). Just extracellular to these interactions, the hydrophobic seal is formed by interactions between residues F43, L46, I47, V50, L95 and L96 (Fig. 2.3a,h,i). Opening a central pore would require substantial conformational rearrangement to break these hydrophobic interactions.

Most ion channels have central pores for conduction, but some, including those of the TMEM16 and OSCA families, have evolved ion pathways through external grooves or tunnels near membrane-facing surfaces[27,28,29,30]. 3a exhibits a distinctive membrane-facing, partially hydrophilic groove is observed between TM2 and TM3 (Fig. 2.3j). It is bounded by the upper tunnel and TM2–TM3 linker at the cytoplasmic and extracellular sides, respectively, and is lined by conserved polar residues N119, N82, S117, Q116, S58, T89, Y113, S92, Y109 and H93 (Fig. 2.3k). The hydrophilic character of this region would be expected to lower the energetic barrier for movement of ions across the outer half of the membrane and more subtle conformational changes would be required to permit permeation along this route relative to a central pore. 3a oligomerization could also be important for building a conduction path either through conformational changes or pore assembly, as was proposed for SARS-CoV-1 3a (ref. [20]). While a pore is not formed in the context of the SARS-CoV-2 tetramer, we observe that one of the partially hydrophilic TM2–TM3 grooves from each dimeric unit face one another across a narrow membrane (Supplemental Figure 2.10c).

2.3.5 Assessment of channel activity in 3a-reconstituted proteoliposomes

We next used electrophysiology and calcium imaging to assess whether SARS-Cov-2 3a is capable of permeating ions. We did not observe consistent currents in recordings from transfected cells and reasoned this could be due to most of 3a being localized to internal membranes (Supplemental Figure 2.11), inactivation at the plasma membrane and/or 3a toxicity. We therefore purified 3a in detergent, reconstituted it into phosphatidylcholine-containing proteoliposomes, and recorded currents across excised patches pulled from proteoliposome blisters. Excised patches from 3a-containing liposomes exhibited currents that reversed at 0 mV and displayed modest outward rectification in symmetric [K^+] (Fig. 2.4a and Supplemental Figure 2.12). In contrast, channel-like activity was not observed in recordings from mock-reconstituted (empty) proteoliposomes (Supplemental Figure 2.12g). We evaluated ion selectivity of 3a by replacing the K^+-containing bath solution with solutions containing Na^+, $NMDG^+$ or Ca^{2+}. Solution exchange resulted in reversal potential shifts from 0.3 ± 0.3 mV in K^+ to -8.2 ± 0.8 mV in Ca^{2+}, -13.5 ± 1.8 mV in Na^+ and -31.0 ± 1.1 mV in $NMDG^+$ (Fig. 2.4b,d). These shifts correspond to the following permeability ratios (P_X/P_{K+}): Ca^{2+} (1.88 ± 0.08) > K^+ (1.0 ± 0.01) > Na^+ (0.59 ± 0.04) > $NMDG^+$ (0.29 ± 0.01) (Fig. 2.4c). Alkaline or acidic pH had little effect on channel activity (Supplemental Figure 2.12f). Consistent with the electrophysiological measurements, we also observed significant calcium influx in 3a proteoliposomes using Fluo-5N calcium imaging following the addition of 8 mM $[Ca^{2+}]_{ext}$ compared to empty liposomes (Fig. 2.4e).

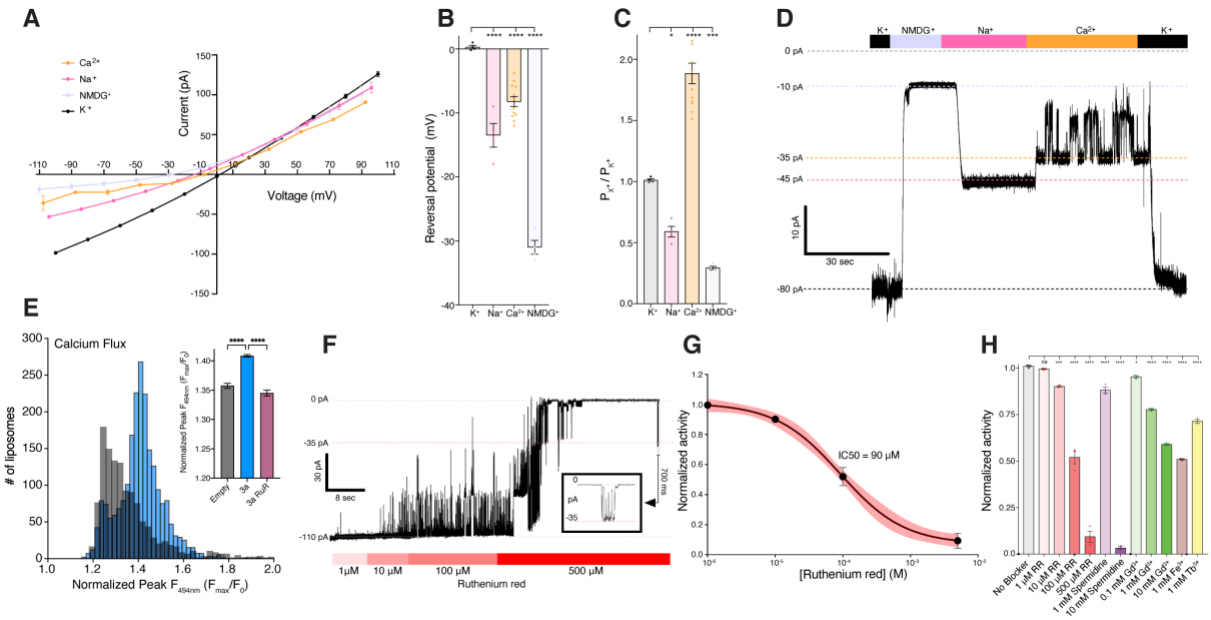


Figure 2.4: Function and inhibition of purified and reconstituted SARS-CoV-2 3a. a, Current–voltage relationship from a 3a-proteoliposome patch. Pipette solution was 150 mM K^+ and external solution was 150 mM K^+ (black), 150 mM Na^+ (pink), 75 mM Ca^{2+} (orange) or 150 mM $NMDG^+$ (blue) (mean ± s.e.m., $n = 3$ recordings from a representative patch). **b**, Reversal potential. Pipette solution was 150 mM K^+ and external solution was 150 mM K^+ (black), 150 mM Na^+ (pink), 75 mM Ca^{2+} (orange) or 150 mM $NMDG^+$ (blue) (mean ± s.e.m., $n = 4$, 4, 12 and 4 patches, respectively). One-way ANOVA with Dunnett correction, ****$P < 0.0001$). **c**, Permeability ratios (P_{X+}/P_{K+}) calculated from **b**. Pipette solution was 150 mM K^+ and external solution was 150 mM K^+ (black), 150 mM Na^+ (pink), 75 mM Ca^{2+} (orange) or 150 mM $NMDG^+$ (blue) (mean ± s.e.m., $n = 4$, 4, 12 and 4 patches, respectively). One-way ANOVA with Dunnett correction, *$P < 0.05$, ***$P < 0.001$, ****$P < 0.0001$). **d**, Gap-free current recording held at −80 mV during bath solution exchanges indicated in the bar above the current trace. **e**, Histogram of peak calcium influx (F_{max}/F) measured by Fluo-5N fluorescence following the addition of 8 mM $[Ca^{2+}]_{ext}$ to 3a liposomes (blue) or empty liposomes (gray). For the inset graph, mean ± s.e.m. of peak calcium influx for 3a liposomes, empty liposomes and 3a liposomes incubated in 200 µM ruthenium red (one-way ANOVA ($F_{(2,4945)} = 99.01$); $n = 1{,}219$, 2,178 and 1,551 liposomes per group, respectively, with Dunnett's multiple-comparisons test, ****$P < 0.0001$). **f**, Gap-free current recording in symmetric 150 mM KCl held at −80 mV during bath solution exchanges of varying ruthenium red concentrations indicated in the bar below the current trace. The boxed inset shows magnified channel openings and closures selected from the region indicated. **g**, Normalized activity in symmetric 150 mM KCl at different concentrations of ruthenium red with fit (black line) and 95% confidence interval (red) shown (mean ± s.e.m., $n = 3$ patches, $IC_{50} = 90 \pm 10$ µM). **h**, Normalized 3a activity in symmetric 150 mM KCl and (from left to right) no blocker, 1 µM ruthenium red ($P = 0.9127$), 10 µM ruthenium red ($P < 0.001$), 100 µM ruthenium red ($P < 0.0001$), 500 µM ruthenium red ($P < 0.0001$), 1 mM spermidine ($P < 0.0001$),, 10 mM spermidine ($P < 0.0001$), 0.1 mM Gd^{3+} ($P = 0.0441$), 1 mM Gd^{3+} ($P < 0.0001$), 10 mM Gd^{3+} ($P < 0.0001$), 1 mM Fe^{3+} ($P < 0.0001$) and 1 mM Tb^{3+} ($P < 0.0001$) (mean ± s.e.m., $n = 3$ patches for each, one-way ANOVA with Dunnett correction, *$P < 0.05$, ***$P < 0.001$, ****$P < 0.0001$). NS, not significant. Source data for **a**–**c**,**e**,**g** and **h** are available online.

We performed mass spectrometry of 3a samples to assess whether channel activity observed in vitro could be due to channel contaminants in purified 3a or lipids used for reconstitution. We analyzed samples of 3a in detergent before liposome incorporation, 3a-reconstituted proteoliposomes and mock-reconstituted liposomes, and searched a database consisting of the 3a construct, the *S. frugiperda* (expression system) proteome, the *Glycine max* (lipid source) proteome and common contaminants (Supplemental Figure 2.13). We detected 3a peptides in both 3a-containing samples, with the detergent sample giving better sequence coverage and more

matched peptides. The only other proteins identified were at low levels and from common contaminants; no peptides from known channels or transmembrane proteins other than 3a were observed in any sample. While we cannot definitively rule out the presence of an unannotated contaminant or one below the detection threshold for mass spectrometry, these results indicate that SARS-CoV-2 3a is responsible for the Ca^{2+}-permeable nonselective channel activity observed in vitro.

2.3.6 Polycations block 3a activity in vitro

We next asked whether blockers of nonselective cation channels inhibit 3a ion conduction in liposomes. We found that ruthenium red, a 786-Da polycationic dye, blocks 3a activity in current recordings (Fig. 2.4g,h) and Ca^{2+} influx measurements (Fig. 2.4e). Ruthenium red displays dose-dependent inhibition of 3a activity in proteoliposome recordings ($IC_{50} = 90 \pm 10$ μM) with flickery block at negative potentials, similar to that observed with other large pore channels (Fig. 2.4g). In addition, we identified the polyamine spermidine, which showed near complete block at 10 mM, and the trivalent ions Gd^{3+}, Fe^{3+} and Tb^{3+}, which showed partial block at 1–10 mM, as 3a inhibitors in proteoliposome recordings (Fig. 2.4g). ruthenium red, spermidine, Gd^{3+} and other trivalents are relatively nonselective inhibitors of cation channels including TRPs, RyRs, CALHMs, K2Ps and K_{IRS}[31,32,33,34,35,36,37], but the spectrum of 3a blockers and affinities observed here is, to our knowledge, distinct from those of other known channels. We further tested reported low-affinity blockers of SARS-CoV-1 3a channel activity[20,21], but found that neither Ba^{2+} nor the small molecule emodin inhibited SARS-CoV-2 3a activity (Supplemental Figure 2.12d,e). Consistently, in a cryo-EM structure of dimeric 3a determined in the presence of 100 μM emodin to 3.7-Å resolution (Supplemental Figures 2.3, 2.6, 2.7, and Table 2.1), we observed no evidence of emodin binding or structural changes compared to apo 3a.

2.3.7 3a permeates the large divalent ion YO-PRO-1

Given the ability of 3a channels to weakly conduct $NMDG^+$ (Fig. 2.4a-d), we next used a YO-PRO-1 fluorescence-based flux assay to assess the ability of 3a channels to conduct other large cations. YO-PRO-1, a 629-Da divalent cation, does not readily cross lipid bilayers and has been used to study the activity of other nonselective cation channels, including P2X7 and TRP channels, that can conduct large organic cations[31,32,38,39]. Robust YO-PRO-1 uptake was observed in a subset of 3a liposomes that was significantly higher than background fluorescence observed in empty liposomes (Fig. 2.5a-d). To compare YO-PRO-1 uptake across multiple experiments, we quantified a single value (area under the curve) to represent both the number of YO-PRO-1+ liposomes and the amount of YO-PRO-1 uptake (Fig. 2.5e). Using this analysis, we found 3a liposomes display significantly greater YO-PRO-1 uptake and accumulation as compared to empty liposomes (Fig. 2.5e). Uptake was dependent on 3a protein because liposomes containing the human K2P channel TRAAK displayed little uptake and accumulation of YO-PRO-1, similar to empty liposomes, as expected for a highly selective K^+ channel (Fig. 2.5a,c-e). Similar to current recordings (Fig. 2.4e,f), we observed a dose-dependent ruthenium red block of 3a activity in the YO-PRO-1 flux assay ($IC_{50} = 175 \pm 98$ μM). These data indicate that 3a can form an ion channel that is capable of passing large cations. Ruthenium red (786 Da) block and YO-PRO-1 (629 Da) and $NMDG^+$ (406 Da) permeation sets limits on 3a pore size that are similar those observed for TRP and P2X channels.

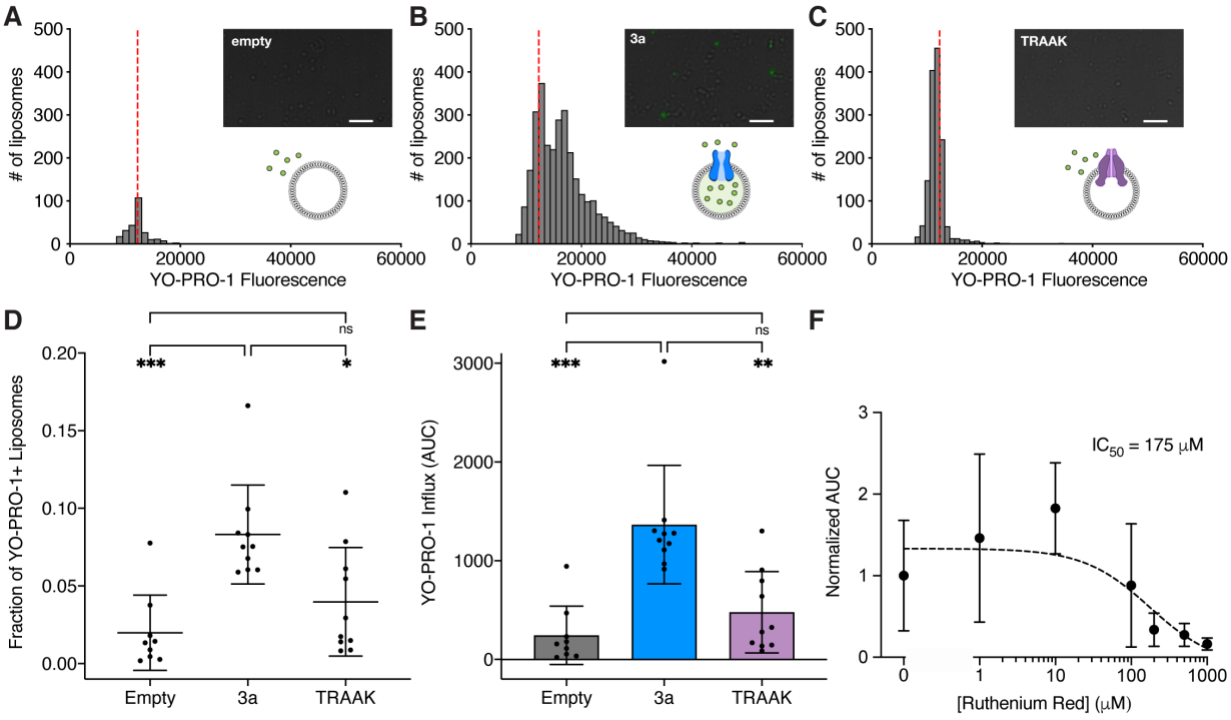


Figure 2.5: A flux assay of SARS-CoV-2 3a channel activity. a–**c**, Histograms from quantified images following a 10-min incubation in 10 μM YO-PRO-1 for empty liposomes (**a**), 3a liposomes (**b**) or TRAAK liposomes (**c**) ($n = 255$, 3,046 and 1,408 for empty, 3a and TRAAK liposomes, respectively). Dotted red lines indicate the position of the mean fluorescence of empty liposomes. Insets, representative images with YO-PRO-1 fluorescence (green) overlaid on brightfield. Scale bars, 20 μm. **d**, Fraction of YO-PRO-1-positive liposomes (mean ± s.d., $n = 9$, 10 and 10 independent replicates for empty, 3a and TRAAK liposomes with 500–5,000 liposomes per replicate, $P = 0.0004$ (one-way ANOVA ($F_{(2,26)} = 10.52$)). Tukey's multiple-comparisons test $P_{\text{empty versus 3a}} = 0.0004$; $P_{\text{TRAAK versus 3a}} = 0.0113$ and $P_{\text{empty versus TRAAK}} = 0.3571$). **e**, Area under the curve (AUC) of YO-PRO-1 fluorescence histograms of the fraction of liposomes that took up YO-PRO-1 per well (mean ± s.d., $n = 9$, 10 and 10 independent replicates for empty, 3a and TRAAK liposomes with 500–5,000 liposomes per replicate, $P < 0.0001$ (Welch's ANOVA test ($W_{(2,16.68)} = 13.22$)), Dunnett's multiple-comparisons test: $P_{\text{empty versus 3a}} = 0.0005$; $P_{\text{TRAAK versus 3a}} = 0.0041$; $P_{\text{empty versus TRAAK}} = 0.4178$). **f**, YO-PRO-1 uptake in 3a liposomes is inhibited by ruthenium red. Mean ± s.d., sigmoidal fit (black dotted line) and IC_{50} are shown. Source data for **a**–**f** are available online.

2.3.8 3a mutants alter ion conduction

We speculate that, relative to the structures presented here, an open conformation of 3a involves substantial rearrangement of the hydrophobic seal to create a central pore or more subtle rearrangement of the partially hydrophilic membrane-facing grooves between TM2 and TM3 to create lateral conduction pathways. To probe the structural determinants for 3a ion conduction, we set out to identify 3a mutations that alter channel activity. To this end, we purified nine mutated or truncated 3a constructs, reconstituted each into proteoliposomes, and compared them to wild-type 3a in electrophysiological recordings (Fig. 2.6a and Supplemental Figure 2.14). We find that two separate mutants alter channel activity, one at the top of the cavity (Q57E) and another and base of the TM2–TM3 grooves (S58L,Q116L). Both Q57E and S58L,Q116L mutants reduce Ca^{2+} and $NMDG^{+}$ permeability without altering Na^{+} or K^{+} permeability. These effects were specific to these mutations as seven additional mutations, including a common circulating variant Q57H[40], had no effect on relative ion permeability. These results are

consistent with the hypothesis that 3a channel opening involves conformational changes of residues that form constrictions above the polar cavity or along the hydrophilic membrane-facing TM2–TM3 grooves.

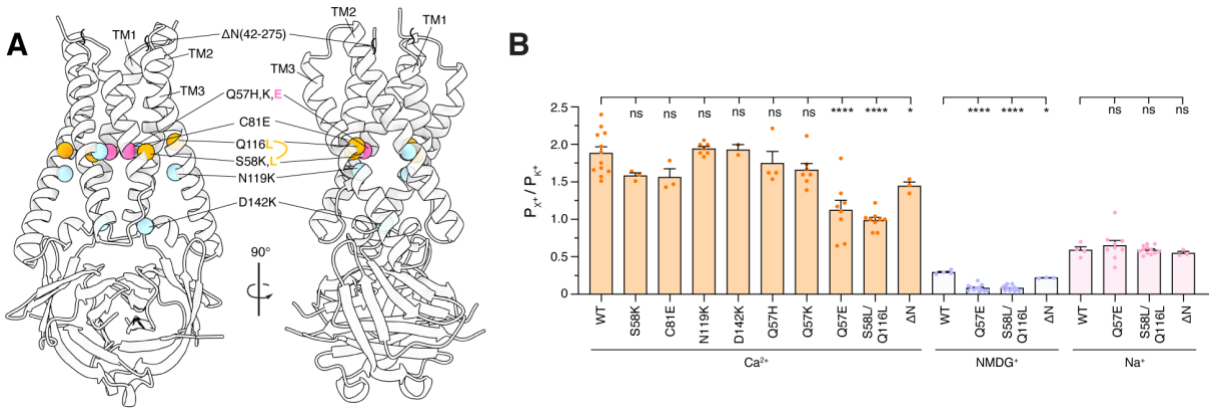


Figure 2.6: Modification of 3a alters channel activity. a, Two views of 3a from the membrane plane with the positions of 3a mutations and truncation indicated. **b**, Permeability ratios (P_X/P_{K+}) calculated from reversal potential shifts for 3a mutations and ΔN truncation (mean ± s.e.m., $n = 8$, 3, 3, 7, 2, 4, 7, 8, 9 and 3 patches for wild-type (WT), S58K ($P = 0.2352$), C81E ($P = 0.1741$), N119K ($P = 0.9971$), D142K ($P = 0.9997$), Q57H ($P = 0.9005$), Q57K ($P = 0.2220$), Q57E ($P < 0.0001$), S58L,Q116L ($P < 0.0001$) and ΔN (for Ca^{2+} $P = 0.0251$ and for $NMDG^+$ $P = 0.0287$), respectively, $^*P < 0.05$, $^{****}P < 0.0001$, one-way ANOVA with Dunnett correction for multiple comparisons). Source data for **b** are available online.

Our structural analysis revealed an unassigned region of density in the cryo-EM maps stretching between subunits just above the extracellular side of the transmembrane region and near the top of the TM2–TM3 grooves that likely corresponds to a portion of the unmodeled N terminus (Supplemental Figure 2.11). To investigate the potential role of the N terminus in channel function, we generated an N-terminal deletion construct lacking the first 41 amino acids (3aΔN). This mutant displayed reduced Ca^{2+} and $NMDG^+$ relative permeability in proteoliposome recordings, albeit to a lesser extent than Q57E and S58L,Q116L (Fig. 2.6b). The relative permeability changes observed in liposomes suggest that residues in the N terminus, such as those at the top of the polar cavity, mediate the ability of 3a to permeate large cations. In addition, the N terminus may be involved in 3a subcellular localization as 3aΔN channels displayed increased localization to the plasma membrane (Supplemental Figure 2.11).

2.3.9 3a-like proteins in Alpha- and Betacoronavirus

While 3a is well conserved in the *Betacoronavirus* subgenus (Supplemental Figure 2.1), related proteins have not yet been identified in other coronaviruses, including the other five species known to infect humans: MERS-CoV, HCoV-NL63, HCoV-229E, HCoV-HKU1 and HCoV-OC43. Thus, we asked whether we could identify more distant homologs using the 3a structure and structure prediction algorithms. 3a homologs were not detected in *Gammacoronavirus*, *Deltacoronavirus* or in the *Betacoronavirus* subgenus *Embecovirus* (which includes HCoV-HKU1 and HCoV-OC43). Distant homology to the CD was identified in the membrane protein ORF5 found in *Betacoronavirus* subgenus *Merbecovirus* species including MERS-CoV. We identified high confidence structural homologs in all remaining *Betacoronavirus* and *Alphacoronavirus* subgenera (including in HCoV-229E and HCoV-NL63 (Supplemental Figure

2.15)), several of which have also been reported to have ion channel activity[41,42,43]. Notably, coronaviruses with 3a homologs have been proposed to derive from viruses that circulate primarily in bats, while coronaviruses without 3a structural homologs have been proposed to derive from viruses that circulate primarily in rodents, birds or pigs (Table 2.2). It may be that the presence of 3a in coronaviruses with bats as their principal reservoir reflects a unique aspect of bat coronavirus biology.

2.4 Discussion

Together, our data demonstrate that SARS-CoV-2 3a adopts a novel dimeric fold widely conserved across Coronaviridae and can form a nonselective Ca^{2+} permeable cation channel. These results provide a basis for further studies to determine mechanisms of 3a channel gating and conduction. A polar cytoplasmic cavity creates a low energy path for ions to traverse between the cytoplasm and halfway across the membrane. However, 3a lacks a central pore across the outer half of the membrane and conformational changes to open one would require breaking interactions that form a hydrophobic seal between TMs in the current structure. Alternatively, lateral conduction pathways could be formed along conserved hydrophilic membrane-facing grooves between TM2 and TM3. Similar lateral conduction paths have precedence among other channels including members of the TMEM16 and OSCA families[27,28,29,30].

Oligomeric 3a assemblies could also mediate 3a activation. 3a dimers associate into tetramers and assembly could feasibly be propagated further into higher-order oligomers. Consistently, we observe increased oligomeric 3a bands in SDS–PAGE gels after liposome incorporation (Supplemental Figures 2.2 and 2.14). This observation, together with the lipid binding pocket within 3a and its localization to organellar membranes, suggests the membrane environment may be important for 3a structure and function.

In liposomes, SARS-CoV-2 3a exhibits permeability to large cations including $NMDG^{+}$ and YO-PRO-1 and sensitivity to ruthenium red and Gd^{3+}, broadly reminiscent of other Ca^{2+}-permeable channels including TRPV1, TRPA1 and P2X7 (refs. [31,32]) while displaying, to our knowledge, a unique functional fingerprint. We note that our data are only partially consistent with previously reported properties of SARS-CoV-1 3a. For example, we do not observe block by Ba^{2+} or emodin and while both proteins apparently form nonselective cation channels, their relative permeability for cations differs. This could be explained by sequence and functional divergence: 3a in the two viruses shares just 73% sequence identity. Alternatively, different recording and lipid environments might alter channel properties. SARS-CoV-1 3a activity was recorded in *Xenopus* oocytes or cultured cells, while SARS-CoV-2 3a activity was recorded in reconstituted membranes. Whether SARS-CoV-2 3a channel activity is observed in cells, perhaps in intracellular membranes where most heterologously expressed protein is localized, remains to be determined.

If 3a has channel activity in cells, it could be important for promoting viral maturation through inhibition of autophagy and disruption of lysosomes[44,45]. The unresolved N and C termini may also play important roles in determining 3a's localization and binding partners. The calcium permeability of 3a is of particular interest in the context of infection. SARS-CoV-2 3a has been

recently shown to trigger programmed cell death in cultured cells[17] and calcium influx through 3a could serve as a switch that activates calcium-dependent caspases and apoptosis. Previous studies showed SARS-CoV-1 3a expression in infected lung pneumocytes, a key cell type also infected by SARS-CoV-2, and calcium signaling in type II pneumocytes plays an important role in maintaining airway homeostasis[46,47]. Thus, the expression of a calcium permeable channel, as for SARS-CoV-2 3a, could affect lung homeostasis and COVID-19 pathogenesis. Our data indicate that 3a could represent a target for treating COVID-19 and other coronavirus diseases and points the way to future experiments to elucidate the role of 3a in the viral life cycle and disease pathology.

2.5 Methods

2.5.1 Cloning and protein expression

The coding sequence for the 3a protein from SARS-Cov-2 was codon optimized for *S. frugiperda* (Sf9 cells) and synthesized (Integrated DNA Technologies). The sequence was then cloned into a custom vector based on the pACEBAC1 backbone (MultiBac, Geneva Biotech) with an added C-terminal PreScission protease (PPX) cleavage site, linker sequence, superfolder GFP (sfGFP) and 7×His tag, generating a construct for expression of 3a-SNS-LEVLFQGP-SRGGSGAAAGSGSGS-sfGFP-GSS-7×His. Mutants and truncation were also introduced into this construct using PCR. MultiBac cells were used to generate a Bacmid according to the manufacturer's instructions. Sf9 cells were cultured in ESF 921 medium (Expression Systems) and P1 virus was generated from cells transfected with Escort IV reagent (MillaporeSigma) according to the manufacturer's instructions. P2 virus was then generated by infecting cells at 2 million cells per ml with P1 virus at a multiplicity of infection of roughly 0.1, with infection monitored by fluorescence and harvested at 72 h. P3 virus was generated in a similar manner to expand the viral stock. The P2 or P3 viral stock was then used to infect Sf9 cells at 4 million cells per ml at a multiplicity of infection of around 2–5. At 72 h, infected cells containing expressed 3a-sfGFP protein were collected by centrifugation at 2,500*g* for 10 min and frozen at −80 °C.

2.5.2 Protein purification

For preparation of the 3a dimer and mutant constructs, infected Sf9 cells from 1 l of culture (roughly 15–20 ml of cell pellets) were thawed in 100 ml of lysis buffer containing 50 mM Tris, 150 mM KCl, 1 mM EDTA pH 8. Protease inhibitors (final concentrations E64 (1 μM), pepstatin A (1 μg ml^{-1}), soy trypsin inhibitor (10 μg ml^{-1}), benzamidine (1 mM), Aprotinin (1 μg ml^{-1}), leupeptin (1 μg ml^{-1}), AEBSF (1 mM) and phenylmethylsulfonyl fluoride (1 mM)) were added to the lysis buffer immediately before use. Benzonase (4 μl) was added after the cell pellet thawed. Cells were then lysed by sonication and centrifuged at 150,000*g* for 45 min. The supernatant was discarded and residual nucleic acid was removed from the top of the membrane pellet using Dulbecco's phosphate buffered saline (DPBS). Membrane pellets were scooped into a dounce homogenizer containing extraction buffer (50 mM tris, 150 mM KCl, 1 mM EDTA, 1% *n*-dodecyl-β-d-maltopyranoside (DDM) (Anatrace), pH 8). A 10% stock solution of DDM was dissolved and clarified by bath sonication in 200 mM tris pH 8 before addition to buffer to the indicated final concentration. Membrane pellets were then homogenized in extraction buffer and

this mixture (150 ml of the final volume) was gently stirred at 4 °C for 1 h. The extraction mixture was centrifuged at 33,000*g* for 45 min and the supernatant, containing solubilized membrane protein, was bound to 4 ml of sepharose resin coupled to anti-GFP nanobody for 1 h at 4 °C. The resin was then collected in a column and washed with 10 ml of buffer 1 (20 mM HEPES, 150 mM KCl, 1 mM EDTA, 0.025% DDM, pH 7.4), 40 ml of buffer 2 (20 mM HEPES, 500 mM KCl, 1 mM EDTA, 0.025% DDM, pH 7.4) and 10 ml of buffer 1. The resin was then resuspended in 6 ml of Buffer 1 with 0.5 mg of PPX protease and rocked gently in the capped column for 2 h. Cleaved 3a protein was then eluted with an additional 8 ml of Wash Buffer, spin concentrated to roughly 500 µl with Amicon Ultra spin concentrator 10-kDa cutoff (Millipore), and then loaded onto a Superdex 200 increase column (GE Healthcare) on an NGC system (Bio-Rad) equilibrated in buffer 1. Peak fractions containing 3a channel were then collected and spin concentrated before incorporation into proteoliposomes or nanodiscs. For the tetramer, the preparation was carried out in a similar manner, except with overnight protease cleavage and collection of a peak of larger hydrodynamic radius (Supplemental Figure 2.2). Traces for gel filtration were prepared using Chromlab 6.0 (Bio-Rad), Prism 8 and Adobe Illustrator.

2.5.3 Proteoliposome formation

For proteoliposome patching experiments, we incorporated protein into lipid and generated proteoliposome blisters for patch recordings using dehydration and rehydration as described previously with the following modifications[48]. The 3a dimer was first purified into buffer 1. Protein was then exchanged into lipid with the addition of Biobeads SM2 (Bio-Rad) and an overnight incubation at a protein:lipid ratio of 1:10 (corresponding to 0.5 mg purified 3a dimer and 5 mg of cleared soybean l-α-phosphatidylcholine, Soy PC, MillaporeSigma) in DR buffer (5 mM HEPES, 200 mM KCl, pH 7.2). For the YO-PRO-1 assay, 3a was incorporated at a ratio of 1:50. TRAAK control proteoliposomes were prepared at 1:50 as described previously[48]. Control liposomes were prepared from the same lipid mix and protocol with protein replaced with buffer 1.

2.5.4 Electrophysiology

All electrophysiology recordings were made from 3a-reconstituted Soy PC proteoliposomes. Patches formed in an inside-out configuration and were quickly (within 5–10 s) transferred to a solution exchange chamber. Recordings were made at room temperature using Clampex v.10.7 data acquisition software (as part of the pClamp v.10.7 suite) with an Axopatch 200B Patch Clamp amplifier and Digidata 1550B digitizer (Molecular Devices) at a bandwidth of 1 kHz and digitized at 500 kHz. A pressure clamp (ALA Scientific) was used to form seals. Potassium pipette and bath solution was 5 mM HEPES pH 7.2 and 150 mM KCl. Sodium bath solution was 5 mM HEPES pH 7.2 and 150 mM NaCl. NaCl in the bath solution was substituted for 150 mM NMDG-Cl or 75 mM $CaCl_2$ for permeability experiments. Borosilicate glass pipettes were pulled and polished to a resistance of 2–5 MΩ when filled with pipette solution. For cation permeability experiments, liquid junction potentials were calculated and data were corrected offline. For current–voltage plots, the following voltage protocol was applied: V_{hold} = 0 mV mV; V_{test} = −100 to +100 mV, Δ20 mV, t_{test} = 1 s. Currents from each patch correspond to mean values during the step to the indicated voltage.

Permeability ratios were calculated according to Goldman–Hodgkin–Katz relationship. For monovalent cations, permeability ratios were calculated as $P_{X+}/P_{K+} = \exp(\Delta V_{rev}F/RT)$. For divalent cations, permeability ratios were calculated as:
$P_{X2+}/P_{K+} = \alpha_{K+}[K +]\exp(\Delta V_{rev}F/RT)(1 + \exp(\Delta V_{rev}F/RT))/ 4\alpha_{X2+}[X^{2+}]$ where V_{rev} is the reversal potential, F is Faraday's constant, R is the universal gas constant, T is absolute temperature (where $RT/F = 25.2$ mV at room temperature) and α is the ion activity coefficient (assumed to be 0.75 for K^+ and 0.25 for Ca^{2+}).

2.5.5 Mass spectrometry

Samples from 3a in DDM (before liposome incorporation), 3a in soy PC (1:50 ratio, and tested for activity in both electrophysiological and fluorescence-based assays) or control liposomes were run on SDS–PAGE gels and stained (Supplemental Figure 2.13). For the DDM sample, the entire gel lane above the dye front was excised, including the stacking gel. For samples with lipid, to minimize lipid interference in sample digestion, the gel lane above the stained lipid was then excised (which includes the region around 15 kDa and larger).

Mass spectrometry was performed by the Vincent J. Coates Proteomics/Mass Spectrometry Laboratory at UC Berkeley. The excised gel lanes were digested with trypsin in situ, and the resulting peptides extracted and concentrated. A nano liquid-chromatography column was packed in a 100-μm inner diameter glass capillary with an emitter tip. The column consisted of 10 cm of Polaris c18 5-μm packing material (Varian). The column was loaded by use of a pressure bomb and washed extensively with buffer A (5% acetonitrile/0.02% heptafluorobutyric acid). The column was then directly coupled to an electrospray ionization source mounted on a Thermo Fisher LTQ XL linear ion trap mass spectrometer. An Agilent 1200 HPLC equipped with a split line so as to deliver a flow rate of 300 nl min^{-1} was used for chromatography. Peptides were separated with a reverse phase elution using a gradient of 5 to 60% acetonitrile generated by a transition from buffer A to buffer B (60% acetonitrile/0.02% heptafluorobutyric acid) over 90 min.

Protein identification was done with Integrated Proteomics Pipeline (IP2, Integrated Proteomics Applications) using ProLuCID/Sequest, DTASelect2 and Census[49,50,51]. Tandem mass spectra were extracted into MS1 and MS2 files from raw files using RawExtractor[52]. Data were searched against a *S. frugiperda* protein database from Uniprot with the added amino acid sequence of purified 3a (including the cleavage site scar), supplemented with sequences of common contaminants and concatenated to form a decoy database[53]. Searches against soy (*G. max*) returned no hits for any sample. LTQ data were searched with 3000.0 milli-amu precursor tolerance and the fragment ions were restricted to a 600.0 ppm tolerance. All searches were parallelized and searched on the VJC proteomics cluster. Search space included all half tryptic peptide candidates with no missed cleavage restrictions. Carbamidomethylation (+57.02146) of cysteine was considered a static modification. To identify authentic termini, we required one tryptic terminus for each peptide identification. The ProLuCID search results were assembled and filtered using the DTASelect program with a peptide false discovery rate (FDR) of 0.001 for single peptides and a peptide FDR of 0.005 for additional peptides for the same protein. Under such filtering conditions, the estimated FDR was less than 1%.

2.5.6 Ca^{2+} uptake assay in proteoliposomes

Fluo-5N was incorporated into 3a-reconstituted proteoliposomes and control liposomes by first thawing frozen liposomes at a ratio of roughly 1:20 (v:v) into modified sucrose formation buffer[54] 25 mM HEPES pH 7.4, 150 mM KCl and 262 mM sucrose with a final concentration of 25 μM Fluo-5N (Stock concentration 5 mM in DMSO, AAT Bioquest). Next, the liposome mixtures in Eppendorf tubes were placed in a foam flotation in an ice bath and sonicated with a Branson Digital Sonifier 450 for 30 s in total (10% amplitude, 10 s pulse with 59-s wait time). Excess (unincorporated) Fluo-5N was then removed using microspin G-50 columns (Cytiva). The sample was then diluted into solution containing 50 mM HEPES, 300 mM KCl and 2 mM EGTA, pH 7.4, plated onto poly-d-lysine (Sigma-Aldrich, 1 mg ml^{-1}) coated 96-well plates and centrifuged (440*g* for 5 min) at room temperature. Ca^{2+} influx was measured on addition of a 50 mM HEPES, 285 mM KCl, 10 mM $CaCl_2$, pH 7.4 solution. Images were acquired every 3 s for a total of 150 s. Fluorescence intensity of each liposome was normalized to its average fluorescence before Ca^{2+} addition.

2.5.7 YO-PRO-1 flux assay

Liposomes were diluted in saline solution containing 300 mM KCl, 50 mM HEPES, pH 7.4 and plated onto poly-d-lysine (Sigma-Aldrich, 1 mg ml^{-1}) coated 96-well plates, centrifuged (5,000*g* for 5 min) and incubated in 10 μM YO-PRO-1 iodide (Invitrogen) for 10 min at room temperature. Liposomes were rinsed with saline solution to remove free YO-PRO-1. Images were acquired using an ImageXpress Micro XLS microscope with a solid-state light source and ×20 air objective. Images were analyzed using MetaXpress 6 software (Molecular Devices). Empty liposomes displayed low-level fluorescence in the 488/540-nm range. Thus, liposomes were defined as YO-PRO-1^+ if their fluorescence intensity was above the average intensity of unlabeled, empty liposomes. To calculate the total number of liposomes, we quantified the total number of liposomes using auto-segmentation on MetaXpress 6 from brightfield images. Stock solutions of ruthenium red (Tocris) were prepared in water, diluted in bath solution and applied to liposomes or cells 10 min before the start of the experiment. All statistical tests were performed using Prism (GraphPad). Values are reported as the mean ± s.d. or mean ± s.e.m as indicated. Mann–Whitney or a one-way or Welch's analysis of variance (ANOVA) followed by the Dunnett's, Sidak's or Tukey's post hoc tests (where appropriate) were used for statistical comparisons.

2.5.8 Nanodisc formation

Freshly purified 3a dimer in Buffer 1 was reconstituted into MSP1E3D1 nanodiscs with a mixture of lipids (DOPE:POPS:POPC at a 2:1:1 mass ratio, Avanti) at a final molar ratio of 1:4:400 (monomer ratio 3a, MSP1E3D1, lipid mixture). First, 20 mM solubilized lipid in Nanodisc formation buffer (20 mM HEPES, 150 mM KCl, 1 mM EDTA pH 7.4) was mixed with additional DDM detergent and 3a protein. This solution was mixed at 4 °C for 30 min before addition of purified MSP1E3D1. This addition brought the final concentrations to approximately 15 μM 3a, 60 μM MSP1E3D1, 6 mM lipid mix and 10 mM DDM in Nanodisc formation buffer. The solution with MSP1E3D1 was mixed at 4 °C for 10 min before addition of 200 mg of Biobeads SM2. Biobeads (washed into methanol, water, and then Nanodisc formation buffer)

were weighed after liquid was removed by pipetting (damp weight). This mix was incubated at 4 °C for 30 min before addition of another 200 mg of Biobeads (for a total 400 mg of Biobeads per 0.5 ml of reaction). This final mixture was then gently tumbled at 4 °C overnight (roughly 12 h). Supernatant was cleared of beads by letting large beads settle and carefully removing liquid with a pipette. Sample was spun for 10 min at 21,000*g* before loading onto a Superdex 200 increase column in 20 mM HEPES, 150 mM KCl, pH 7.4. Peak fractions corresponding to 3a protein in MSP1E3D1 were collected, 10-kDa cutoff spin concentrated and used for grid preparation. MSP1E3D1 was prepared as described[55] without cleavage of the His-tag. Tetrameric 3a in nanodiscs was prepared similarly, except with a ratio of 1:2:200 (monomer ratio 3a, MSP1E3D1, lipid mixture).

2.5.9 Grid preparation

Dimeric 3a in MSP1E3D1 was prepared at final concentration of 1.1 mg ml^{-1}. For the sample with emodin (MillaporeSigma, cat. no. E7881), a stock solution of 50 mM emodin in DMSO added to protein sample for final concentrations of 1.1 mg ml^{-1} 3a and 100 µM emodin and 1% DMSO. Concentrated sample was cleared by a 10 min 21,000*g* spin at 4 °C before grid making. For freezing grids, a 3-µl drop of protein was applied to freshly glow discharged Holey Carbon, 300 mesh R 1.2/1.3 gold grids (Quantifoil). A FEI Vitrobot Mark IV (Thermo Fisher Scientific) was used with 4 °C, 100% humidity, 1 blot force, a wait time of around 5 s and a 3 s blot time, before plunge freezing in liquid ethane. Grids were then clipped and used for data collection. Tetrameric 3a in MSP1E3D1 was frozen at 0.7 mg ml^{-1} with the same grid preparation. Unimaged grids from the same session were then shipped to the Netherlands for data collection on the 300 kV microscope with Falcon 4/CFEG/Selectris energy filter.

2.5.10 Cryo-EM data acquisition

For data collected on the Talos Arctica, grids were clipped and transferred to the microscope operated at 200 kV. Fifty frame videos were recorded on a Gatan K3 Summit direct electron detector in super-resolution counting mode with pixel size of 0.5685 Å. For the apo 3a dataset, the electron dose was 9.528 and 10.135 e^{-} $Å^2$ s^{-1} and total dose was 50.02 and 53.72 e^{-} $Å^2$ in the first set (1–2,007) and second set (2,008–6,309) of videos, respectively. The two different doses are the result of needing to restart the electron gun during collection. For the 3a with added emodin dataset, the electron dose was 8.991 e^{-} $Å^2$ s^{-1} and total dose was 47.21 e^{-} $Å^2$. For the 3a tetramer, the electron dose was 8.841 e^{-} $Å^2$ s^{-1} and total dose was 49.95 e^{-} $Å^2$. Nine videos were collected around a central hole position with image shift and defocus was varied from −0.6 to −2.0 µm through SerialEM[56].

For data collected for the high-resolution 3a dimer structure, clipped grids from the same batch used on Talos Artica were sent to Thermo Fisher Scientific RnD division in Eindhoven, Netherlands. Grids were loaded onto the Krios G4 microscope equipped cold field emission gun (CFEG) operated at 300 kV. Data were collected on a Falcon 4 detector that was mounted behind a Selectris X energy filter. The slit width of the energy filter was set to 10 eV. Then 5,599 video stacks containing 1,429 raw frames were collected with the electron event representation (EER) mode[57] of Falcon 4 detector at a magnification of ×165,000 corresponding to a pixel size of

0.727 Å. Each video stack was recorded with an exposure time of 6 s with a total dose of 50 e^{-} /Å^2 on sample and a defocus range between 0.5 and 1.2 μm.

Also see Table 2.1 for detailed data collection statistics.

2.5.11 Cryo-EM data processing

For the apo 3a dimer, motion correction and dose weighting were performed on all 6,309 videos using RELION 3.1's implementation of MotionCor2, and twice 'binned' to 1.137 Å per pixel[58,59,60]. CTFFIND-4.1 was used to estimate the contrast transfer function (CTF) parameters[61]. Micrographs were then manually sorted to eliminate subjectively bad micrographs, such as empty or contaminated holes, resulting in 3,611 good micrographs. Additionally, micrographs with a CTF maximum resolution lower than 4 Å were discarded, resulting in 2,595 remaining micrographs. Template-free auto-picking of particles was performed with RELION3.1's Laplacian-of-Gaussian filter yielding an initial set of particles. This initial set of particles were iteratively classified to generate templates, which were subsequently used to template-based auto-pick 1,750,730 particles.

Template-picked particles were iteratively two-dimensionally classified in RELION3.1 and then in cryoSPARC v.2 (ref. [62]), resulting in 820,543 particles. These particles were subsequently three-dimensionally classified in cryoSPARC v.2 with iterative ab initio and heterogeneous refinement jobs. The resulting maps were visually evaluated with regard to the transmembrane domain density. A set of 86,479 particles were identified, polished in RELION3.1 and refined in cryoSPARC v.2 with subsequent homogeneous and nonuniform refinement[63] jobs (maps were low-pass filtered to an initial resolution where TM density was still visible (6–9 Å), and the dynamic mask was tightened with the near (2–5 Å) and far (3–9 Å) parameters), yielding a map with overall resolution of 3.6 Å. UCSF pyem tools were used to convert data from cryoSPARC to RELION format[64].

From this set of 86,479 particles, 2D classification was performed in RELION3.1 to identify a set of particles with subjectively equal view distribution. From the resulting set, 1,000 particles were randomly sampled and their coordinates used for training in the Topaz particle-picking pipeline[65]. Training, picking and extraction were performed independently on each subset of the micrographs. Then, 4,134,279 total particles were extracted in RELION3.1 with a box size of 256 pixels and 'binned' four times to 4.548 Å per pixel. These particles were then iteratively two-dimensionally classified in RELION3.1 resulting in 2,674,606 particles that were extracted at 2.274 Å per pixel. 2D classification was continued in both RELION3.1 and cryoSPARC v.2 resulting in 1,429,763 particles. Further classification was performed in cryoSPARC v.2 with subsequent ab initio (four classes, maximum resolution 8 Å) and heterogeneous refinement (8 Å initial resolution) jobs. The two best classes were selected and the particles pooled resulting in 743,800 particles, which were extracted in RELION3.1 at 1.137 Å per pixel.

Iterative three-dimensional (3D) classification was performed with subsequent ab initio and heterogeneous refinement jobs as described above. Following each round, 2D classification jobs were used to 'rescue' good particles from the worst classes before the next round. After three rounds, a final 2D classification job was used to identify 112,502 particles, which were

subsequently pooled with the previous 86,479 RELION3.1 template-picked particles, resulting in 185,871 particles after duplicates (within 100 Å) were removed with RELION3.1.

These particles were then refined with subsequent homogeneous and nonuniform refinement jobs resulting in a map with overall resolution of 3.4 Å. This map was postprocessed in RELION3.1 using a mask with a soft edge (5-pixel extension, 7-pixel soft edge), the output of which was used for Bayesian particle polishing in RELION3.1 (training and polishing were each performed independently on each subset of the micrographs). The resulting 'shiny' particles were then refined in cryoSPARC v.2 with subsequent homogenous refinement (one extra pass, 7 Å initial resolution) and nonuniform refinement (C2, 1 extra pass, 9 Å initial resolution) to yield a map with 2.9-Å overall resolution.

For the 3a dimer with added 100 μM emodin, initial processing was similar to the dimer without the added drug (Supplemental Figure 2.7). As with the apo 3a dimer, the critical steps included Topaz particle picking, particle clean-up with cryoSPARC v.2 ab initio and heterogeneous refinement, nonuniform refinement with tightened masking and RELION3.1 Bayesian particle polishing. However, in contrast to the apo dataset, we observed a set of particles that were included in <4-Å reconstructions that had discontinuous transmembrane domain density. Removal of these particles with RELION3.1 3D classification without angular sampling led to the best map from the emodin-added dataset. We did not see any evidence of bound emodin, but the 1% DMSO added with drug addition may have contributed to subtle map differences (Supplemental Figures 2.6 and 2.7).

For the 3a tetramer, the initial 7,092 micrographs were first cleaned using manual inspection and removal of images with <4 Å CtfMaxResolution to obtain a set of 4,324 micrographs. Reference particles for Topaz particle picking were generated by first template picking in RELION3.1, followed by 2D classification in both RELION3.1 and cryoSPARC v.2, and subsequent ab initio in cryoSPARC v.2. Particles from various views were then selected from iterative RELION3.1 2D classification to create a set of 6,843 particles. Using these coordinates for training, Topaz particle picking was then performed to generate a set of 1,282,913 initial particles. These particles were then cleaned using 2D classification in RELION3.1 and cryoSPARC v.2, followed by rounds of cryoSPARC v.2 ab initio and RELION3.1 3D classification. A major hurdle for tetramer processing was obtaining a reconstruction where most particles were properly oriented in the same direction (that is, intracellular domains on the same side of the nanodisc as seen in the 2D classes, Fig. 2.2f). Substantial, clean-up by 3D classification was needed to generate a correctly aligned reference map, but this map could then be used as a reference for refinements and classification for larger particle sets. Reconstructions with C1 or C2 symmetry looked similar (Supplemental Figure 2.3e), although no tetramer reconstruction went to high enough resolution to determine symmetry with certainty. Therefore, it is possible that either the tetramer is pseudosymmetric or that different particles have heterogeneous orientations between dimer pairs. For the tetramer, the highest resolution reconstruction came from cryoSPARC v.2 nonuniform refinement with a tightened mask, which was subsequently used for dimer-docking and figure preparation.

For the apo 3a dimer imaged on a Krios with CFEG, Selectris X and Falcon 4, the EER video motion correction and subsequent polishing was performed in RELION3.1 using the devel-eer

branch of Relion. Dose fractions consisting of 30 frames corresponding to 1.035 e^{-} /Å^2 per fraction were created. The initial 5,599 micrographs were pruned by selecting 4,495 micrographs with <3.5 Å CtfMaxResolution that also passed manual inspection. Similar to the Arctica datasets, an initial set of particles was generated with template-based picking and subsequent 2D classification, ab initio and heterogeneous refinement for clean-up was performed in cryoSPARC v.2 (Supplemental Figure 2.9). This particle set (44,944) was then used as training for the first round of Topaz particle picking. For initial particle clean-up, the most important step was heterogeneous refinement in cryoSPARC v.2, yielding a particle set (215,227) that gave a C2 nonuniform refinement at 2.69 Å. For final particle clean-up, using RELION3.1, polished particles were subjected to 3D classification with no angular sampling and high tau (40) during which we monitored the rounds of classification by looking for convergence in resolution and inspecting maps to check for quality of protein features or noise. This type of job reliably allowed us to select the best particles (61,531) for high-resolution reconstruction at 2.26 Å.

At this point, a second Topaz training and picking round was conducted using this set of particles. After merging the best particles from both rounds of Topaz picking, removing potential duplicate particles within a distance of 100 Å and subsequent processing, we obtained a set of particles (91,218) that achieved a nonuniform reconstruction at 2.17 Å. Finally, rounds of CtfRefine in cryoSPARC v.2 and a final nonuniform refinement (with higher-order CTF terms enabled) achieved a map with an estimated resolution of 2.1 Å. Depending on inputs into sharpening and local resolution jobs (Supplemental Figure 2.8a), the map has regions that give resolution estimates below 2 Å, consistent with our observations of holes in aromatic side chain density.

We note that the merging of two Topaz picked particle sets likely allowed us to break a particle-limited barrier to achieve the final reconstruction based on ResLog[66] analysis performed in cryoSPARC (Supplemental Figure 2.8c). Finally, we sampled various box sizes during processing (as large as 416 pixels and as small as 256 pixels)[67]. The best reconstructions were consistently achieved with our final box size of 300 pixels (218.1 Å).

2.5.12 Modeling, refinement, and analysis

Apo dimeric 3a cryo-EM maps were sharpened using cryoSPARC and were of sufficient quality for de novo model building in Coot[68]. Real space refinement of the models was carried out using Phenix.real_space_refine[69]. Molprobity[70] was used to evaluate the stereochemistry and geometry of the structure for subsequent rounds of manual adjustment in Coot and refinement in Phenix. For final sharpening and visualization of the high-resolution map we used Phenix Resolve Density Modification[71]. Docking of the apo dimeric 3a into the tetrameric 3a cryo-EM map was performed in Phenix using a map in which large empty regions of the nanodisc were erased in Chimera[72]. Similar results were found using maps with only the CDs present. Cavity measurements were made with HOLE[73] implemented in Coot. Comparisons to the structure database was performed with Dali[26]. Structure prediction was performed with Phyre2 (ref. [74]). Figures were prepared using PyMOL, Chimera, ChimeraX[75], Fiji, Prism 8, GNU Image Manipulation Program and Adobe Photoshop and Illustrator software.

2.5.13 Fluorescence size-exclusion chromatography

Sf9 cells (roughly 4 million) from the third day of infection were pelleted, frozen and then thawed into extraction buffer (20 mM Tris pH 8, 150 mM KCl, all protease inhibitors used for protein purification, 1 mM EDTA, 1% DDM). Extraction was performed at 4 °C for 1 h and lysate was then pelleted at 21,000*g* at 4 °C for 1 h to clear supernatant. Supernatant was then run on a Superose 6 Increase column with fluorescence detection for GFP into 20 mM HEPES pH 7.4, 150 mM KCl and 0.025% DDM.

2.5.14 Transfection for confocal imaging

The constructs for full length 3a and 3aΔN were cloned into a vector with a CMV-promoter and C-terminal enhanced green fluorescent protein. Constructs (2 μg) were transfected into human embryonic kidney 293 cells on glass coverslips using Fugene HD (Promega) per the manufacturer's instructions. Two days after transfection, cells were washed with DPBS and then fixed in 4% formaldehyde in DPBS for 10 min. Cells were then washed with DPBS before mounting the coverslip with Prolong Glass Antifade with NucBlue (Thermo Fisher Scientific) per the manufacturer's instructions. Fluorescent images were collected using a Zeiss LSM 880 NLO AxioExaminer confocal microscope at either ×20 (numerical aperture 1.0) or ×63 oil immersion objective (numerical aperture 1.4). The samples were excited with 488 nm argon laser and we carried out image analysis using ImageJ (v.1.53a).

2.6 Data availability

All data associated with this study are publicly available. For cryo-EM of dimeric apo 3a, the final model is in the PDB under 6XDC, the final map is in the Electron Microscopy Data Bank (EMDB) under EMD-22136, and the original micrograph videos and final particle stack is in the Electron Microscopy Public Image Archive (EMPIAR) under EMPIAR-10439. For tetrameric apo 3a, the final map is in the EMDB under EMD-22138, and the original micrograph videos and final particle stack is in EMPIAR under EMPIAR-10441. For dimeric 3a in the presence of emodin, the final map is in the EMDB under EMD-22139, and the original micrograph videos and final particle stack is in EMPIAR under EMPIAR-10440. For the high-resolution dimeric apo 3a, the final model is in the PDB under 7KJR, the final map is in the EMDB under EMD-22898, and the original micrograph videos and final particle stack is in EMPIAR under EMPIAR-10612. For mass spectrometry searches, the *S. frugiperda* proteome was downloaded from Uniprot database (https://www.uniprot.org/uniprot/?query=taxonomy:7108). Source data are provided with this paper (https://www.nature.com/articles/s41594-021-00619-0).

2.7 Acknowledgements

We thank H. Adesnik for providing emodin and for discussions. We thank P. Tobias for computational resources at the Cal-Cryo-EM facility and J. Hurley and E. Nogales for supporting the microscopy work. We thank Thermo Fisher Scientific for microscope access. We thank L. Kohlstaedt for assistance with mass spectrometry. We thank R. Rietmeijer for his gift of TRAAK proteoliposomes. We thank members of the Brohawn laboratory and A. Noble for thoughtful feedback on our preprint. Finally, we thank the many people at UC Berkeley and surrounding companies working during the pandemic who helped make this project possible. This work used the Vincent J Proteomics/Mass Spectrometry Laboratory at UC Berkeley, supported in part by

NIH S10 Instrumentation grant no. S10RR025622. S.G.B. is a New York Stem Cell Foundation-Robertson Neuroscience Investigator. This work was funded by the New York Stem Cell Foundation; NIGMS grant no. GM123496; a McKnight Foundation Scholar Award; a Rose Hill Innovator Award; a Sloan Research Fellowship (to S.G.B.); NIGMS grant no. GM128263 (to D.M.K.); NSF Graduate Research Fellowship DGE1752814 (to S.S.M.); a Howard Hughes Medical Institute Faculty Scholar Award (to D.M.B.); and a Fast Grants Award from Emergent Ventures at the Mercatus Center, George Mason University (to D.M.B., S.G.B. and H. Adesnik).

2.8 Contributions

B.S., D.M.K., S.S.M., A.K., D.M.B. and S.G.B. conceived the project. D.M.K. performed all molecular biology, biochemistry, preparation of proteoliposomes and cryo-EM sample preparation. B.S. performed all electrophysiology. S.S.M. designed and performed the YO-PRO-1, ruthenium red blocker and Ca^{2+} flux assays with help from D.M.B. C.M.H. and D.M.K. processed the cryo-EM data. S.S. performed light microscopy. J.P.R. and D.B.T. collected cryo-EM data at UC Berkeley, and A.K. collected cryo-EM data at Thermo Fisher Scientific. S.G.B. built and refined the atomic models. D.M.B. and S.G.B. secured funding and supervised research. D.M.K., B.S., C.M.H., S.S.M., A.K., D.M.B. and S.G.B. wrote the paper with input from all authors.

2.9 Competing interests

A.K. is an employee of Thermo Fisher Scientific. The other authors declare no competing interests.

2.10 References

1. Wrapp, D. et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020).
2. Walls, A. C. et al. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **183**, 1735 (2020).
3. Shang, J. et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221–224 (2020).
4. Zhang, L. et al. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **368**, 409–412 (2020).
5. Yin, W. et al. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* **368**, 1499–1504 (2020).
6. Zhang, Y. & Kutateladze, T. G. Molecular structure analyses suggest strategies to therapeutically target SARS-CoV-2. *Nat. Commun.* **11**, 2920–2924 (2020).
7. Gordon, D. E. et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).
8. Zhong, X. et al. Amino terminus of the SARS coronavirus protein 3a elicits strong, potentially protective humoral responses in infected patients. *J. Gen. Virol.* **87**, 369–373 (2006).

9. Lu, B. et al. Humoral and cellular immune responses induced by 3a DNA vaccines against severe acute respiratory syndrome (SARS) or SARS-like coronavirus in mice. *Clin. Vaccin. Immunol.* **16**, 73–77 (2009).
10. Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nat. Med.* **26**, 450–452 (2020).
11. Channappanavar, R., Fett, C., Zhao, J., Meyerholz, D. K. & Perlman, S. Virus-specific memory CD8 T cells provide substantial protection from lethal severe acute respiratory syndrome coronavirus infection. *J. Virol.* **88**, 11034–11044 (2014).
12. Wang, H. et al. SARS-CoV-2 proteome microarray for mapping COVID-19 antibody interactions at amino acid resolution. *ACS Cent. Sci.* **6**, 2238–2249 (2020).
13. Ward, D. et al. An integrated in silico immuno-genetic analytical platform provides insights into COVID-19 serological and vaccine targets. *Genome Med* **13**, 4–12 (2021).
14. Siu, K.-L. et al. Severe acute respiratory syndrome coronavirus ORF3a protein activates the NLRP3 inflammasome by promoting TRAF3-dependent ubiquitination of ASC. *FASEB J.* **33**, 8865–8877 (2019).
15. Chan, C.-M. et al. The ion channel activity of the SARS-coronavirus 3a protein is linked to its pro-apoptotic function. *Int. J. Biochem. Cell Biol.* **41**, 2232–2239 (2009).
16. Yue, Y. et al. SARS-coronavirus open reading frame-3a drives multimodal necrotic cell death. *Cell Death Dis.* **9**, 904–915 (2018).
17. Ren, Y. et al. The ORF3a protein of SARS-CoV-2 induces apoptosis in cells. *Cell Mol. Immunol.* **17**, 881–883 (2020).
18. Castaño-Rodriguez, C. et al. Role of severe acute respiratory syndrome coronavirus viroporins E, 3a, and 8a in replication and pathogenesis. *Mbio* **9**, 439 (2018).
19. Silvas, J. et al. Contribution of SARS-CoV-2 accessory proteins to viral pathogenicity in K18 hACE2 transgenic mice. Preprint at *bioRxiv* https://doi.org/10.1101/2021.03.09.434696 (2021).
20. Lu, W. et al. Severe acute respiratory syndrome-associated coronavirus 3a protein forms an ion channel and modulates virus release. *Proc. Natl Acad. Sci. USA* **103**, 12540–12545 (2006).
21. Schwarz, S., Wang, K., Yu, W., Sun, B. & Schwarz, W. Emodin inhibits current through SARS-associated coronavirus 3a protein. *Antivir. Res.* **90**, 64–69 (2011).
22. Delcour, A. H. (ed.) *Electrophysiology of Unconventional Channels and Pores*. (Springer International Publishing, 2015).
23. Nieva, J. L., Madan, V. & Carrasco, L. Viroporins: structure and biological functions. *Nat. Rev. Microbiol.* **10**, 563–574 (2012).
24. Scott, C. & Griffin, S. Viroporins: structure, function and potential as antiviral targets. *J. Gen. Virol.* **96**, 2000–2027 (2015).
25. Nakane, T. et al. Single-particle cryo-EM at atomic resolution. *Nature* **587**, 152–156 (2020).
26. Holm, L. & Sander, C. Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480 (1995).
27. Dang, S. et al. Cryo-EM structures of the TMEM16A calcium-activated chloride channel. *Nature* **552**, 426–429 (2017).
28. Paulino, C., Kalienkova, V., Lam, A. K. M., Neldner, Y. & Dutzler, R. Activation mechanism of the calcium-activated chloride channel TMEM16A revealed by cryo-EM. *Nature* **552**, 421–425 (2017).

29. Zhang, M. et al. Structure of the mechanosensitive OSCA channels. *Nat. Struct. Mol. Biol.* **25**, 850–858 (2018).
30. Jojoa-Cruz, S. et al. Cryo-EM structure of the mechanically activated ion channel OSCA1.2. *Elife* **7**, e41845 (2018).
31. Chung, M.-K., Güler, A. D. & Caterina, M. J. TRPV1 shows dynamic ionic selectivity during agonist stimulation. *Nat. Neurosci.* **11**, 555–564 (2008).
32. Banke, T. G., Chaplan, S. R. & Wickenden, A. D. Dynamic changes in the TRPA1 selectivity filter lead to progressive but reversible pore dilation. *Am. J. Physiol. Cell Physiol.* **298**, C1457–C1468 (2010).
33. Xu, L., Tripathy, A., Pasek, D. A. & Meissner, G. Ruthenium red modifies the cardiac and skeletal muscle Ca^{2+} release channels (ryanodine receptors) by multiple mechanisms. *J. Biol. Chem.* **274**, 32680–32691 (1999).
34. Choi, W., Clemente, N., Sun, W., Du, J. & Lu, W. The structures and gating mechanism of human calcium homeostasis modulator 2. *Nature* **576**, 163–167 (2019).
35. Pope, L., Lolicato, M. & Minor, D. L. Polynuclear ruthenium amines inhibit K2P channels via a 'finger in the dam' mechanism. *Cell Chem. Biol.* **27**, 511–524.e4 (2020).
36. Xie, L.-H., John, S. A. & Weiss, J. N. Spermine block of the strong inward rectifier potassium channel Kir2.1: dual roles of surface charge screening and pore block. *J. Gen. Physiol.* **120**, 53–66 (2002).
37. Suma, A., Granata, D., Thomson, A. S., Carnevale, V. & Rothberg, B. S. Polyamine blockade and binding energetics in the MthK potassium channel. *J. Gen. Physiol.* **152**, 6726 (2020).
38. Surprenant, A., Rassendren, F., Kawashima, E., North, R. A. & Buell, G. The cytolytic P2Z receptor for extracellular ATP identified as a P2X receptor (P2X7). *Science* **272**, 735–738 (1996).
39. Karasawa, A., Michalski, K., Mikhelzon, P. & Kawate, T. The P2X7 receptor forms a dye-permeable pore independent of its intracellular domain but dependent on membrane lipid composition. *Elife* **6**, 3393 (2017).
40. Hadfield, J. et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).
41. Müller, M. A. et al. Human Coronavirus NL63 Open Reading Frame 3 encodes a virion-incorporated N-glycosylated membrane protein. *Virol. J.* **7**, 6 (2010).
42. Wang, K. et al. PEDV ORF3 encodes an ion channel protein and regulates virus production. *FEBS Lett.* **586**, 384–391 (2012).
43. Zhang, R. et al. The ORF4a protein of human coronavirus 229E functions as a viroporin that regulates viral production. *Biochim. Biophys. Acta Biomembr.* **1838**, 1088–1095 (2014).
44. Miao, G. et al. ORF3a of the COVID-19 virus SARS-CoV-2 blocks HOPS complex-mediated assembly of the SNARE complex required for autolysosome formation. *Dev. Cell* **56**, 427–442.e5 (2020).
45. Ghosh, S. et al. β-coronaviruses use lysosomes for egress instead of the biosynthetic secretory pathway. *Cell* **183**, 1520–1535.e14 (2020).
46. Zhao, J., Zhao, J. & Perlman, S. T cell responses are required for protection from clinical disease and for virus clearance in severe acute respiratory syndrome coronavirus-infected mice. *J. Virol.* **84**, 9318–9325 (2010).

47. Dietl, P., Haller, T. & Frick, M. Spatio-temporal aspects, pathways and actions of Ca^{2+} in surfactant secreting pulmonary alveolar type II pneumocytes. *Cell Calcium* **52**, 296–302 (2012).
48. del Mármol, J., Rietmeijer, R. A. & Brohawn, S. G. in *Potassium Channels: Methods and Protocols* (eds Shyng, S.-L. et al.) 129–150 (Springer, 2018).
49. Xu, T. et al. ProLuCID: an improved SEQUEST-like algorithm with enhanced sensitivity and specificity. *J. Proteom.* **129**, 16–24 (2015).
50. Tabb, D. L., McDonald, W. H. & Yates, J. R. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* **1**, 21–26 (2002).
51. Park, S. K., Venable, J. D., Xu, T. & Yates, J. R. A quantitative analysis software tool for mass spectrometry–based proteomics. *Nat. Methods* **5**, 319–322 (2008).
52. McDonald, W. H. et al. MS1, MS2, and SQT—three unified, compact, and easily parsed file formats for the storage of shotgun proteomic spectra and identifications. *Rapid Commun. Mass Spectrom.* **18**, 2162–2168 (2004).
53. Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J. & Gygi, S. P. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC–MS/MS) for large-scale protein analysis: the yeast proteome. *J. Proteome Res.* **2**, 43–50 (2003).
54. Schöneberg, J. et al. ATP-dependent force generation and membrane scission by ESCRT-III and Vps4. *Science* **362**, 1423–1428 (2018).
55. Ritchie, T. K. et al. Chapter 11—Reconstitution of membrane proteins in phospholipid bilayer nanodiscs. *Methods Enzymol.* **464**, 211–231 (2009).
56. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
57. Guo, H. et al. Electron-event representation data enable efficient cryoEM file storage with full preservation of spatial and temporal resolution. *IUCrJ* **7**, 860–869 (2020).
58. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
59. Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5–17 (2019).
60. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, 163 (2018).
61. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
62. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
63. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 (2020).
64. Asarnow, D., Palovcak, E. & Cheng, Y. asarnow/pyem: UCSF pyem v.0.5 (Zenodo, 2019); https://doi.org/10.5281/zenodo.3576630
65. Bepler, T. et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019).
66. Stagg, S. M., Noble, A. J., Spilman, M. & Chapman, M. S. ResLog plots as an empirical metric of the quality of cryo-EM reconstructions. *J. Struct. Biol.* **185**, 418–426 (2014).

67. Herzik, M. A., Wu, M. & Lander, G. C. High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM. *Nat. Commun.* **10**, 1032–1039 (2019).
68. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
69. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877 (2019).
70. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
71. Terwilliger, T. C., Ludtke, S. J., Read, R. J., Adams, P. D. & Afonine, P. V. Improvement of cryo-EM maps by density modification. *Nat. Methods* **17**, 923–927 (2020).
72. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
73. Smart, O. S., Neduvelil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph* **14**, 354–360 (1996).
74. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
75. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

2.11 Tables

Data collection	**Dimeric apo 3a (2.9 Å)**	**Dimeric 3a with emodin**	**Tetrameric apo 3a**	**Dimeric apo 3a (2.1 Å)**
PDB	6XDC	n/a	n/a	7KJR
EMDB	22136	22139	22138	22898
EMPIAR	10439	10440	10441	10612
Total movies #	6309	6750	7092	5598
Selected movies #	2595	3405	4324	4495
Magnification	36,000 x	36,000 x	36,000 x	165,000 x
Voltage (KV)	200	200	200	300
Electron exposure ($e^-/Å^2$)	50.325 or 53.72 (1-2007) (2008-6309)	47.21	49.95	50
Frame #	50	50	50	1449 or 1379 (EER)
Defocus range (um)	-0.6 to -2.0	-0.6 to -2.0	-0.6 to -2.0	-0.5 to -1.1
Super resolution pixel size ($Å^2$)	0.5685	0.5685	0.5685	0.3685
Binned pixel size ($Å^2$)	1.137	1.137	1.137	0.727
Processing				
Initial particle images (no.)	4,134,279	3,873,767	1,282,913	2,314,293
Final particle images (no.)	185,871	51,908	64,410	91,218
Map resolution Masked (Å, FSC = 0.143)	2.9	3.69	6.5	2.08

Symmetry imposed	C2	C2	C2	C2
Refinement				
Model resolution (Å, FSC = 0.143 / FSC = 0.5)	3.2/3.6			2.0 / 2.2
Map-sharpening B factor (Å^2)	-111.2			-43.5
Composition				
Number of atoms	3150			3840
Number of protein residues	386			448
Number of ligands				2
Number of waters				122
R.m.s. deviations				
Bond lengths (Å)	0.006			0.005
Bond angles (Å)	0.785			0.729
Validation				
MolProbity score	1.55			1.42
Clashscore	4.63			7.60
Ramachandran plot				
Favored (%)	96.03			98.17
Allowed (%)	3.97			1.13
Disallowed (%)	0			0
Rotamer outliers (%)	1.15			0.25
Mean B factor (Å^2)				
Protein	108.61			23.30
Ligand				51.72
Water				22.11

Table 2.1: Cryo-EM data collection, refinement and validation statistics

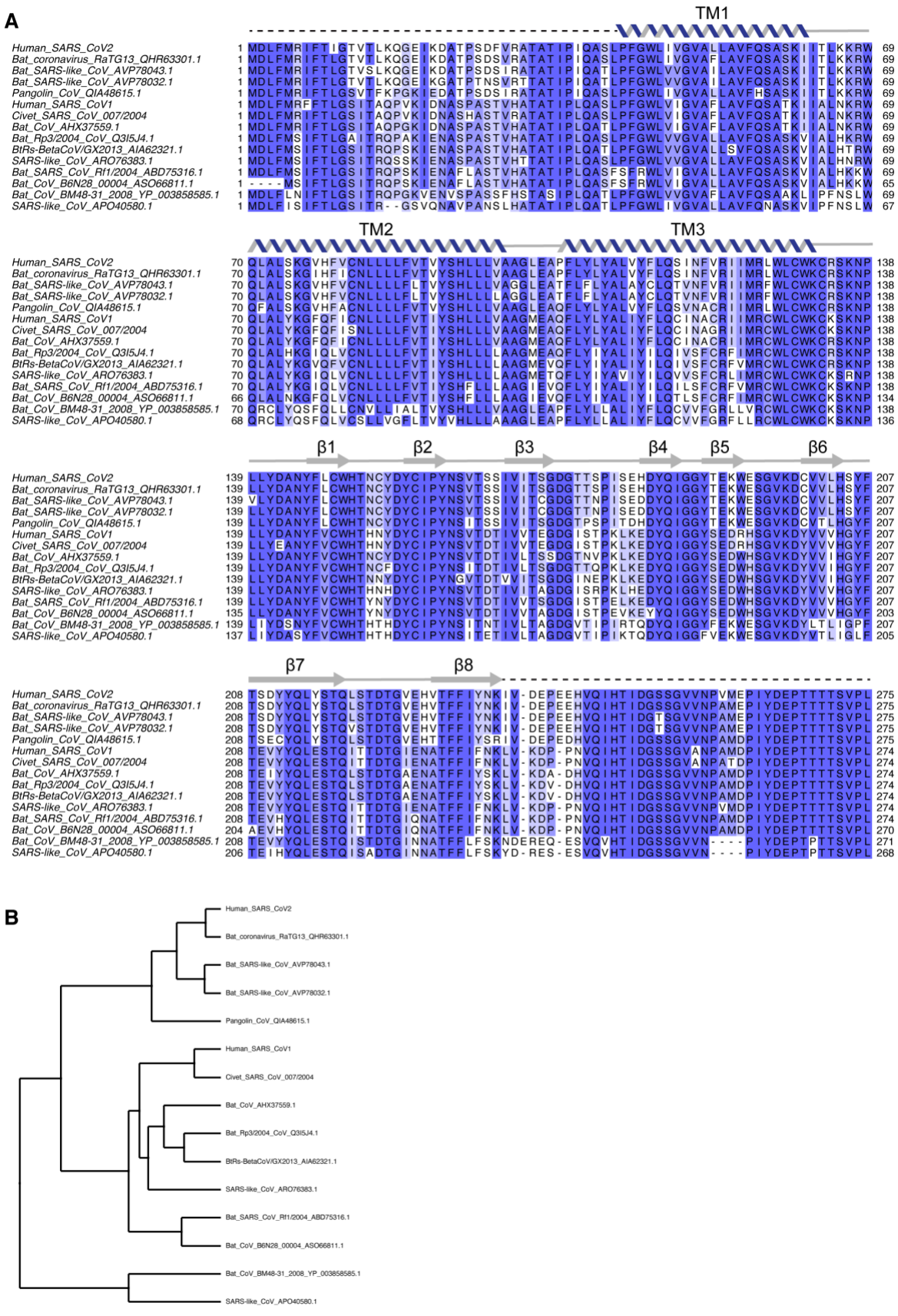
Proposed host origin		Subgenus / Species	Annotated ORF name	Accession	Phyre predicted amino acid range	Phyre confidence score
Bat	Alphacoronavirus	Colacovirus				
		Bat coronavirus CDPHE15 (1384461)	NS3	YP_008439203.1	69 - 222	98%
		Decacovirus				
		Bat coronavirus HKU10 (1244203)	NS3	AFU92105.1	69 - 221	96%
		R. ferrumequinum HuB-2013 (2501926)	ORF3	YP_009199791.1	69 - 220	97%
		Duvinacovirus				
		Human coronavirus 229E (11137)	ORF4	ARU07602.1	69 - 221	95%
		Luchacovirus				
		Coronavirus AcCoV-JC34 (1964806)	ORF3	YP_009380522.1	69 - 220	96%

		Minacovirus				
		Mink coronavirus 1 (766791)	NS3	YP_009019183.1	68-221	96%
		Minunacovirus				
		Miniopterus bat coronavirus 1 (694000)	ORF3	ACA52165.1	69 - 221	96%
		Miniopterus bat coronavirus HKU8 (694001)	ORF3	AIA62228.1	69 - 221	96%
		Myotacovirus				
		Myotis ricketti alphacoronavirus Sax-2011 (2501927)	ORF3	AIA62247.1	78-211	97%
		Nyctacovirus				
		Nyctalus velutinus alphacoronavirus SC-2013 (2501928)	ORF3	YP_009201731.1	69 - 220	97%
		Pipistrellus kuhlii coronavirus 3398 (2492658)	ORF3	YP_009755891.1	69 - 220	97%
		Pedacovirus				
		Porcine epidemic diarrhea virus (28295)	ORF3	AWM99571.1	69 - 221	96%
		Scotophilus bat coronavirus 512 (693999)	ORF3	YP_001351685.1	69 - 220	97%
		Rhinacovirus				
		Rhinolophus bat coronavirus HKU2 (693998)	ORF3	ATN23890.1	69 - 220	97%
		Setracovirus				
		Human coronavirus NL63 (277944)	ORF3	AGT51388.1	69 - 221	97%
		NL63-related bat coronavirus BtKYNL63-9b (1920748)	ORF3	YP_009824968.1	69 - 220	96%
		Tegacovirus				
		Alphacoronavirus 1 (693997)	NSP3B	AEM55568.1	69 - 220	96%
	Betacoronavirus	**Sarbecovirus**				
		Human SARS-CoV-1 (694009)	ORF3A	P59632	42-237	100%
		Human SARS-CoV-2 (2697049)	ORF3A	YP_009724391.1	n/a	n/a
		Hibecovirus				
		Bat Hp-betacoronavirus Zhejiang2013 (1541205)	ORF3	YP_009072441.1	42-236	100%
		Zaria bat coronavirus (989337)	ORF3	ADY17912.1	101-233	100%
		Nobecovirus				
		Eidolon bat coronavirus C704 (983924)	ORF3	ADX59467.1	123-190	96%
		Rousettus bat coronavirus GCCDC1 (1892416)	NS3	YP_009273006.1	122-209	95%
		Rousettus bat coronavirus HKU9 (694006)	NS3	QJX58367.1	122-209	95%
		Merbecovirus				
		Human Middle East respiratory syndrome-related (1335626)	ORF5	QJX19961.1	119-146	88%
		Hedgehog coronavirus 1 (1965093)	ORF5	QCC20718.1	126-205	86%
		Pipistrellus bat coronavirus HKU5 (694008)	ORF5	AWH65914.1	126-182	92%

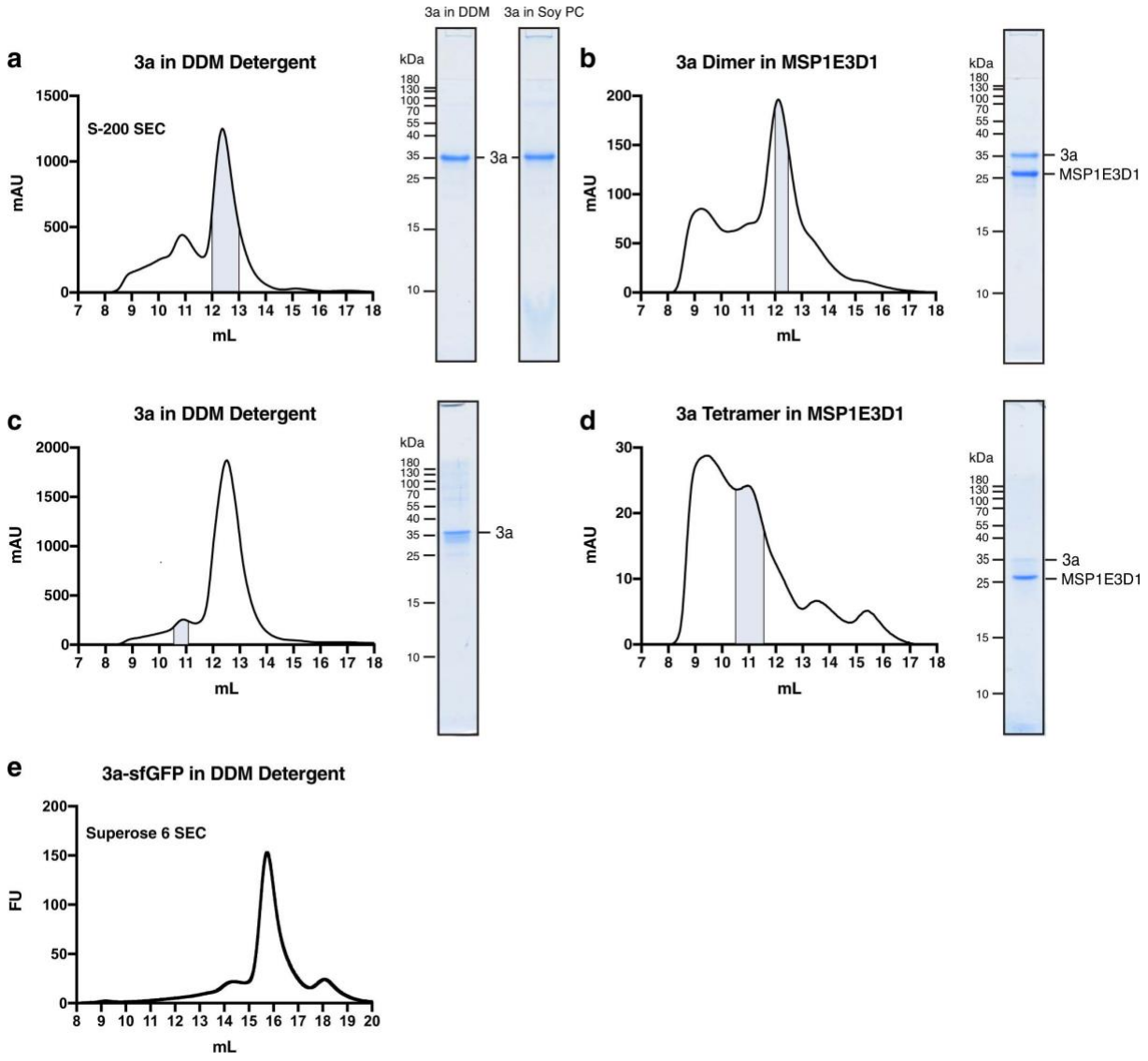
		Tylonycteris bat coronavirus HKU4 (694007)	ORF5	AWH65903.1	126-182	91%
Rodent		Embecovirus				
		Betacoronavirus 1 (694003)	none	none	none	none
		China Rattus coronavirus HKU24 (2501960)	none	none	none	none
		Human coronavirus HKU1 (290028)	none	none	none	none
		Murine coronavirus (694005)	none	none	none	none
		Myodes coronavirus 2JL14	none	none	none	none
Avian / Pig	Deltacoronavirus	Andecovirus				
		Wigeon coronavirus HKU20 (1159908)	none	none	none	none
		Buldecovirus				
		Bulbul coronavirus HKU11 (574549)	none	none	none	none
		Common moorhen coronavirus HKU21	none	none	none	none
		Coronavirus HKU15 (1965089)	none	none	none	none
		Munia coronavirus HKU13 (1297661)	none	none	none	none
		White-eye coronavirus HKU16 (1159907)	none	none	none	none
		Herdecovirus				
		Night heron coronavirus HKU19	none	none	none	none
	Gammacoronavirus	Brangacovirus				
		Goose coronavirus CB17	none	none	none	none
		Cegacovirus				
		Beluga whale coronavirus SW1 (694015)	none	none	none	none
		Igacovirus				
		Avian coronavirus (694014)	none	none	none	none
		Avian coronavirus 9203	none	none	none	None
		Duck coronavirus 2714 (300188)	none	none	none	None

Table 2.2: 3a homologs across *Coronaviridae*

2.12 Supplemental figures

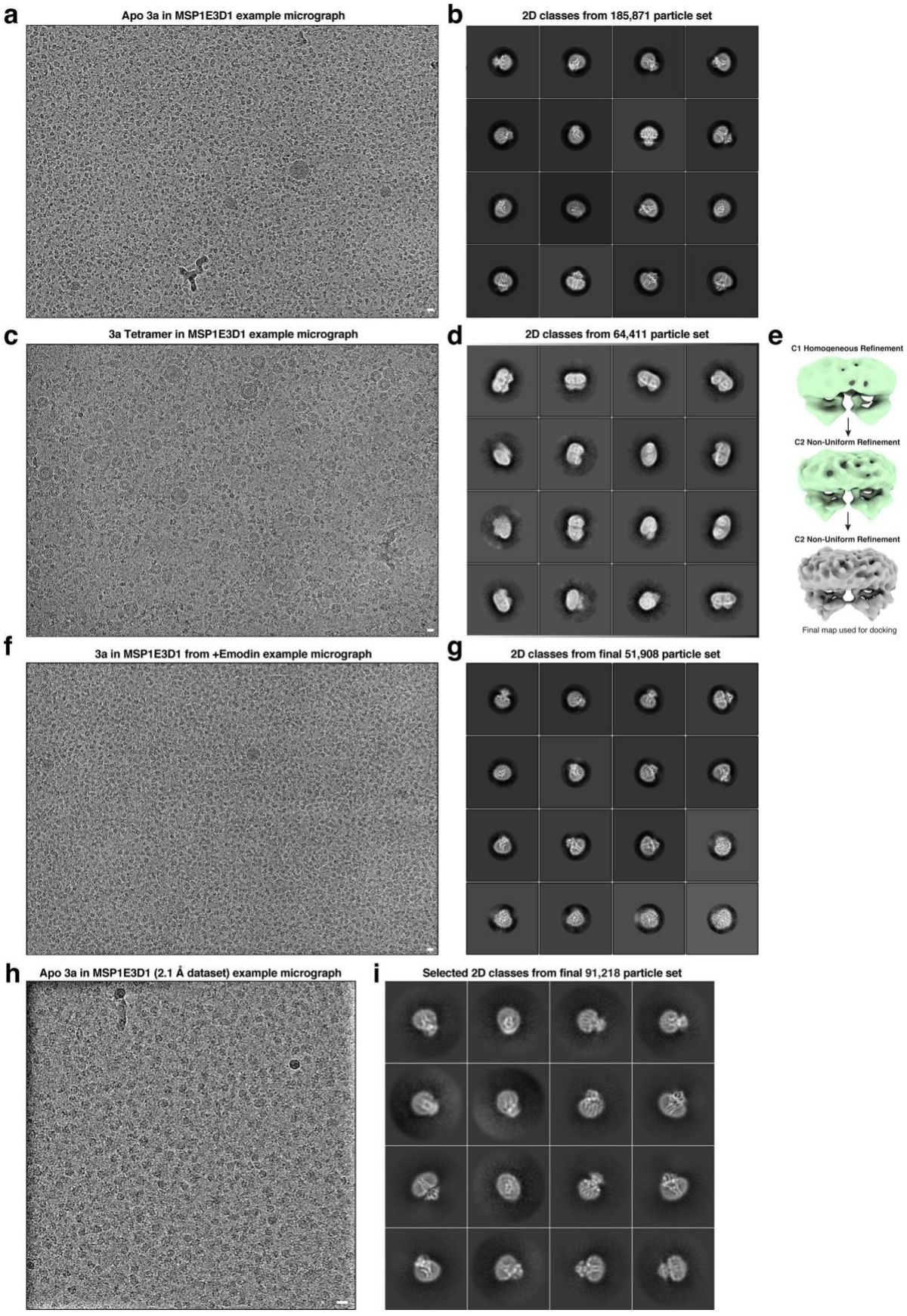

A
TM1
TM2
TM3
β1
β2
β3
β4
β5
β6
β7
β8
Human_SARS_CoV2
Bat_coronavirus_RaTG13_QHR63301.1
Bat_SARS-like_CoV_AVP78043.1
Bat_SARS-like_CoV_AVP78032.1
Pangolin_CoV_QIA48615.1
Human_SARS_CoV1
Civet_SARS_CoV_007/2004
Bat_CoV_AHX37559.1
Bat_Rp3/2004_CoV_Q3I5J4.1
BtRs-BetaCoV/GX2013_AIA62321.1
SARS-like_CoV_ARO76383.1
Bat_SARS_CoV_Rf1/2004_ABD75316.1
Bat_CoV_B6N28_00004_ASO66811.1
Bat_CoV_BM48-31_2008_YP_003858585.1
SARS-like_CoV_APO40580.1
B
Human_SARS_CoV2
Bat_coronavirus_RaTG13_QHR63301.1
Bat_SARS-like_CoV_AVP78043.1
Bat_SARS-like_CoV_AVP78032.1
Pangolin_CoV_QIA48615.1
Human_SARS_CoV1
Civet_SARS_CoV_007/2004
Bat_CoV_AHX37559.1
Bat_Rp3/2004_CoV_Q3I5J4.1
BtRs-BetaCoV/GX2013_AIA62321.1
SARS-like_CoV_ARO76383.1
Bat_SARS_CoV_Rf1/2004_ABD75316.1
Bat_CoV_B6N28_00004_ASO66811.1
Bat_CoV_BM48-31_2008_YP_003858585.1
SARS-like_CoV_APO40580.1

Supplemental Figure 2.1: Sequence alignment of 3a from Betacoronavirus Sarbecovirus a, Alignment of fifteen 3a protein sequences colored by conservation in a ramp from white (not conserved) to dark blue (highly conserved). Accession numbers are indicated. Sequences were selected to maximize diversity among annotated Sarbecovirus 3a proteins. Secondary structure for SARS-CoV-2 is drawn above the sequence with unmodeled sequence drawn as dashed lines. **b,** Neighbor-joining tree calculated from the alignment in (**a**).

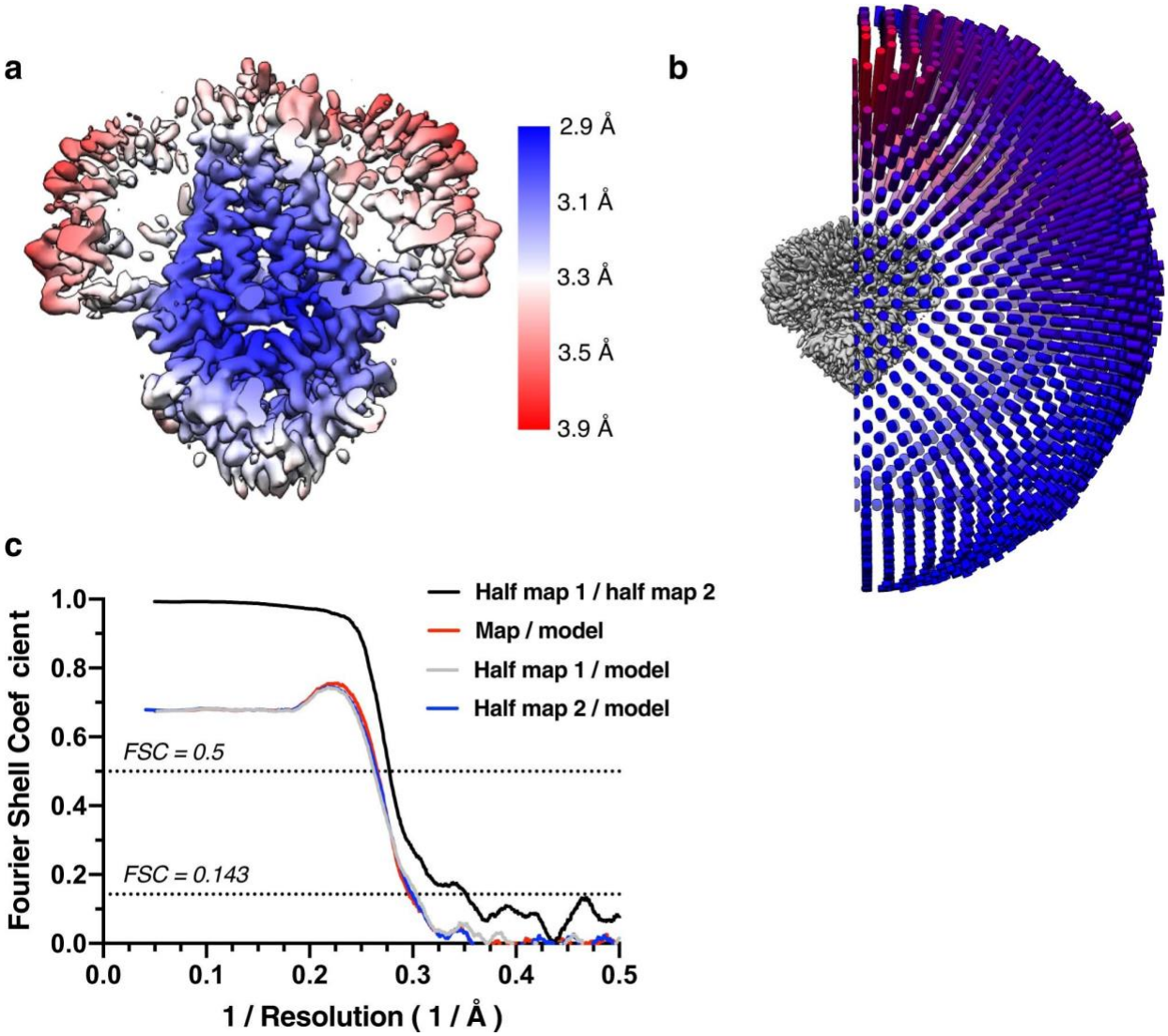


Supplemental Figure 2.2: Purification and reconstitution of 3a. a, Size exclusion chromatogram of 3a expressed in insect cells and extracted and purified in DDM (left). Pooled fractions corresponding to dimeric 3a are highlighted in blue. Coomassie-stained SDS-PAGE of pooled dimeric 3a-containing fractions (center) and of 3a following reconstitution into PC lipids (right). This is a representative preparation used for 3a proteoliposome experiments. WT 3a was purified into DDM six times with similar biochemical behavior. Proteoliposomes from three Soy PC preparations were used for activity assays. **b**, Size exclusion chromatogram of dimeric 3a reconstituted into MSP1E3D1 lipid nanodiscs (left). Pooled fractions are highlighted blue. 3a was incorporated into nanodiscs three separate times with similar biochemical behavior, and this representative preparation was used for dimeric 3a cryo-EM data collection (**c**,**d**) Same as (**a**,**b**), but for tetrameric 3a. The tetrameric peak has been observed in all WT 3a preparations. This representative preparation was used for cryo-EM of tetrameric 3a. **e**, GFP fluorescence

chromatogram of 3a expressed in SF9 cells and extracted in DDM detergent. Samples were run on a Superose 6 column.


a
Apo 3a in MSP1E3D1 example micrograph
b
2D classes from 185,871 particle set
c
3a Tetramer in MSP1E3D1 example micrograph
d
2D classes from 64,411 particle set
e
C1 Homogeneous Refinement
C2 Non-Uniform Refinement
C2 Non-Uniform Refinement
Final map used for docking
f
3a in MSP1E3D1 from +Emodin example micrograph
g
2D classes from final 51,908 particle set
h
Apo 3a in MSP1E3D1 (2.1 Å dataset) example micrograph
i
Selected 2D classes from final 91,218 particle set

Supplemental Figure 2.3: Example micrographs and 2D class averages of 3a structures. a, Representative micrograph and (**b**) 2D class averages of dimeric apo 3a in MSP1E3D1 lipid nanodiscs. **c**, Representative micrograph and (**d**) 2D class averages of tetrameric apo 3a in MSP1E3D1 lipid nanodiscs. **e**, Map overview pipeline for final steps of tetramer processing (Also see Methods) (**f**) Representative micrograph and (**g**) 2D class averages of dimeric 3a in MSP1E3D1 lipid nanodiscs with emodin added. **h**, Representative micrograph and (**i**) 2D class averages of the Krios/CFEG/Selectris collected dimeric apo 3a in MSP1E3D1 lipid nanodiscs (from cryoSPARC). Micrograph scale bars, 100 Å.



Supplemental Figure 2.4: Cryo-EM validation for dimeric apo 3a in MSP1E3D1 lipid nanodiscs. a, Local resolution estimated in Relion colored as indicated on the final map. **b**, Angular distribution of particles used in final refinement with final map for reference. **c**, Fourier Shell Correlation (FSC) relationships (unmasked) between (black) the two unfiltered half-maps from refinement and used for calculating overall resolution at 0.143, (red) the final map and model, (gray) half-map one and model, and (blue) half-map and model.

Template picked 1,750,730 initial particles,
86,479 final particles

Topaz picked 4,134,279 particles, 4.548 Å/pixel

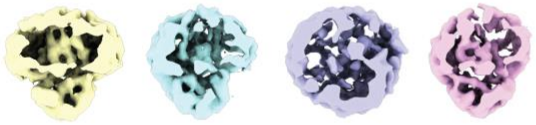
Iterative 2D classification

2,674,606 particles

Extract at 2.274 Å/pixel
Iterative 2D classification

1,429,763 particles

Ab-initio



383,318 particles | 354,561 particles | 347,736 particles | 344,148 particles

Heterogeneous refinement

401,642 particles | 338,227 particles | 342,158 particles

743,800 particles

Extract at 1.137 Å / pixel

Continued at top right

Continued from left

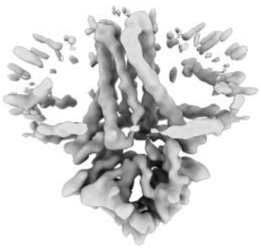
3 rounds:
Ab-initio
Heterogeneous refinement
2D classification

112,502 particles

Pool with 86,479 template-pick particles
Remove duplicates

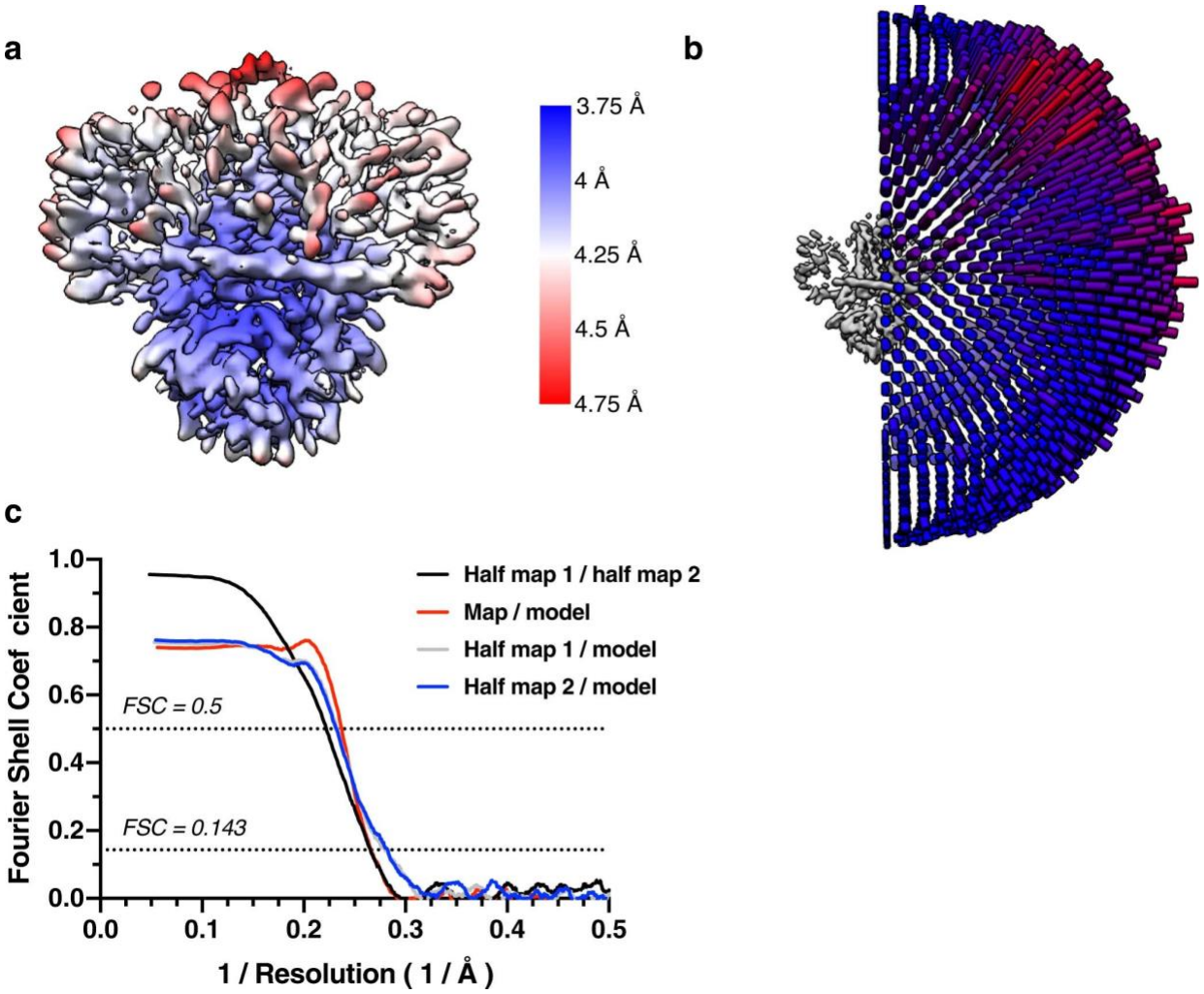
185,871 particles

Homogeneous refinement
Non-uniform refinement (C2)



Bayesian particle polishing
Homogeneous refinement
Non-uniform refinement (C2)

Supplemental Figure 2.5: Cryo-EM processing pipeline for dimeric apo 3a in MSP1E3D1 lipid nanodiscs. Overview of Cryo-EM data processing pipeline in cryoSPARC and Relion. See Methods for details.



Supplemental Figure 2.6: Cryo-EM validation for dimeric 3a in MSP1E3D1 lipid nanodiscs with emodin added. a, Local resolution estimated in Relion colored as indicated on the final map. **b**, Angular distribution of particles used in final refinement with final map for reference. **c**, Fourier Shell Correlation (FSC) relationships (masked) between (black) the two unfiltered half-maps from refinement and used for calculating overall resolution at 0.143, (red) the final map and model, (gray) half-map one and model, and (blue) half-map and model.

3a in MSP1E3D1 + Emodin Dataset

Micrograph clean-up (Start: 6750 mics)
Select, CtfMaxResolution sub-4 Å resolution (3676 mics)
Manual inspection (3405 mics)

For Topaz input:
5678 training particles from partial dataset processing using manual template picking for references (particles chosen from a range of 2D classed particle views)

Topaz picked particles (3,873,767), 4.548 Å/pixel

↓ Iterative 2D classification (Relion)

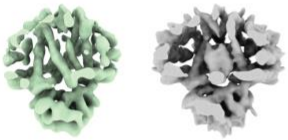
3,121,137 particles, 2.274 Å/pixel

↓ Iterative 2D classification (Cryosparc)

1,895,307 particles

↓ Ab initio for good and bad templates
Iterative Heterogeneous Refine
Until no improvement in resolution

256,290 particles, re-extract to 1.137 Å/pixel



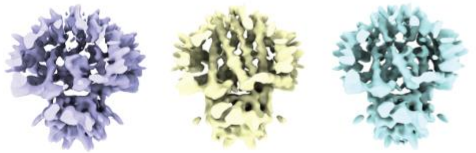
Reference

Non-uniform refinement A
Starting resolution 10 Å
Mask near 3 Å , Far 6 Å
Reference: gaussian filtered lower resolution prior NU

↓ Relion Refine3D
Reference from cryosparc NU
Soft mask around protein made in Relion
Output used for Bayesian Polishing

Non-uniform refinement B
Settings: As Non-uniform refinement A with polished particles

↓ Relion Refine3D
Reference from cryosparc NU
Mask from Cryosparc NU (refine_mask)



Relion Class3D
3 Classes, Tau 16, No alignments
Angles from previous Relion Refine3D
Reference from previous cryosparc NU
Mask from previews Cryosparc NU (refine_mask)

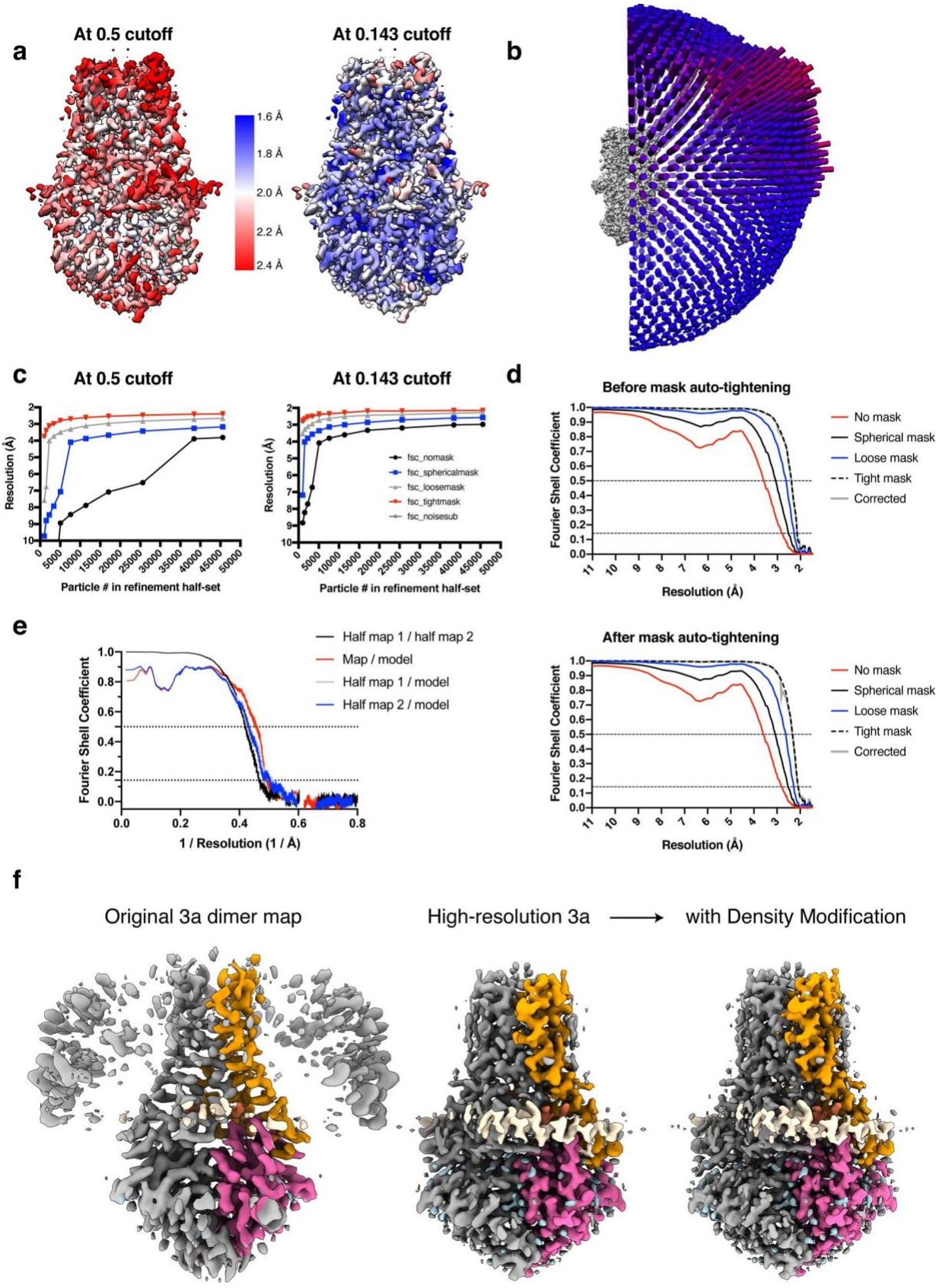
↓ 51,908 particles (continuous TM density)

↘ 143,375 particles (discontinuous TM density)

Non-uniform refinement C and D
Settings: Non-uniform refinement A

Final Map
Used for post-processing
and map and model comparison

Supplemental Figure 2.7: Cryo-EM processing pipeline for dimeric 3a in MSP1E3D1 lipid nanodiscs with emodin added. Overview of Cryo-EM data processing pipeline in cryoSPARC and Relion. See Methods for details.


a
At 0.5 cutoff
At 0.143 cutoff
1.6 Å
1.8 Å
2.0 Å
2.2 Å
2.4 Å
b
c
At 0.5 cutoff
At 0.143 cutoff
Resolution (Å)
Particle # in refinement half-set
fsc_nomask
fsc_sphericalmask
fsc_loosemask
fsc_tightmask
fsc_noisesub
d
Before mask auto-tightening
Fourier Shell Coefficient
Resolution (Å)
No mask
Spherical mask
Loose mask
Tight mask
Corrected
e
Half map 1 / half map 2
Map / model
Half map 1 / model
Half map 2 / model
1 / Resolution (1 / Å)
After mask auto-tightening
f
Original 3a dimer map
High-resolution 3a
with Density Modification

Supplemental Figure 2.8: Cryo-EM validation for 2.1 Å dimeric apo 3a in MSP1E3D1 lipid nanodiscs. a, Local resolution estimated in cryoSPARC at the indicated FSC thresholds colored as indicated on the density modified map. **b**, Angular distribution of particles used in final refinement with map for reference. **c**, ResLog analysis conducted in cryoSPARC at the indicated FSC thresholds. **d**, Fourier Shell Correlation as calculated in cryoSPARC before (top) and after (bottom) mask auto-tightening in the final round of refinement. **e**, Fourier Shell Correlation (FSC) relationships (masked) calculated in Phenix between (black) the two unfiltered half-maps from refinement and used for calculating overall resolution at 0.143, (red) the final map and model, (gray) half-map one and model, and (blue) half-map and model. **f**, Comparison of the original (2.9 Å) 3a map left to the high-resolution (2.1 Å) map before (middle) and after (right) Phenix density modification.

Apo 3a dimer in MSP1E3D1 (2.1 Å dataset)

Micrograph clean-up (Start: 5598 mics)
CtfMaxResolution sub-3.5 Å (4663 mics)
Manual inspection (4495 mics)

Start Particles: 2,314,293
(Topaz picked - Duplicates)
Final Particles: 91,218

For Topaz input round #1:
44,944 training particles from partial dataset processing using manual template picking

For Topaz input round #2:
61,531 training particles from topaz pick processing round #1

Topaz round #1 (1,707,274), Box 256 px, 2.908 Å/px

2D classification (Relion)

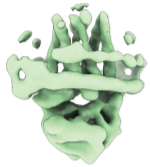
1,625,956 particles, Box 256 px, 1.454 Å/px

2x Ab Initio (cryoSPARC)
2 Class, C1 symmetry

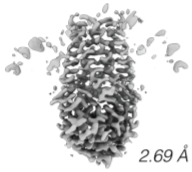
925,324 particles

3x Heterogeneous Refine (cryoSPARC)
3 Class, C2 symmetry

215,227 particles, Box 256 px, 0.727 Å/px (physical px)



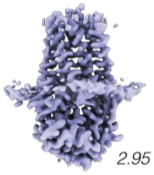
Reference



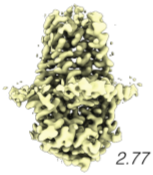
2.69 Å

Non-uniform refinement
Initial lowpass resolution 6 Å
Mask near 3 Å , Far 6 Å
C2 symmetry
Reference from Het Refine Class

Relion Refine3D
Reference from cryoSPARC NU
Soft mask from cryoSPARC NU
Bayesian Polishing in Relion
CtfRefinement in Relion



2.95 Å



2.77 Å

Relion Class3D
2 Classes, Tau 40, No angular sampling
Angles from previous Relion Refine3D
Reference from Refine3D
Mask from cryoSPARC NU (refine_mask)

61,531 particles

2.26 Å

Non-uniform refinement
Initial lowpass resolution 6 Å
Mask near 6 Å , Far 14 Å
C2 symmetry
Reference from previous NU

Topaz round #2 (2,245,142), Box 256 px, 1.454 Å/px

4x Heterogeneous Refine (cryoSPARC)
3 Class, C2 symmetry
References from Round #1

300,615 particles, Box 256 px, 1.454 Å/px

JoinStar with original 61,531 template particles and duplicate removal at 50 Å

305,331 particles, Box 320 px, 0.727 Å/px

Ab Initio, C1 symmetry, 2 class

191,299 particles

Processing performed similarly to topaz round #1

68,008 particles

2.33 Å

Merged particles from round 1 and 2
Removed Duplicates at 100 Å distance

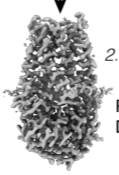
91,799 particles, 300 px box 0.727 Å/px (Re-Extracted)

NU Refinement (*2.31 Å*), Ab Initio reference
Refine3D
Bayesian Polishing
2D Class in cryoSPARC

91,218 particles

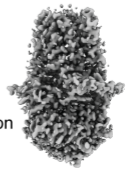
NU Refinement (*2.17 Å*)
CtfRefine-NU Refine-CtfRefine in cryoSPARC (Max fit 1 Å)

Non-uniform refinement
Initial lowpass resolution 6 Å
Mask near 6 Å , Far 14 Å
C2 symmetry
Reference from ab initio (6 Å)
Higher Order CTF enabled
One extra final pass

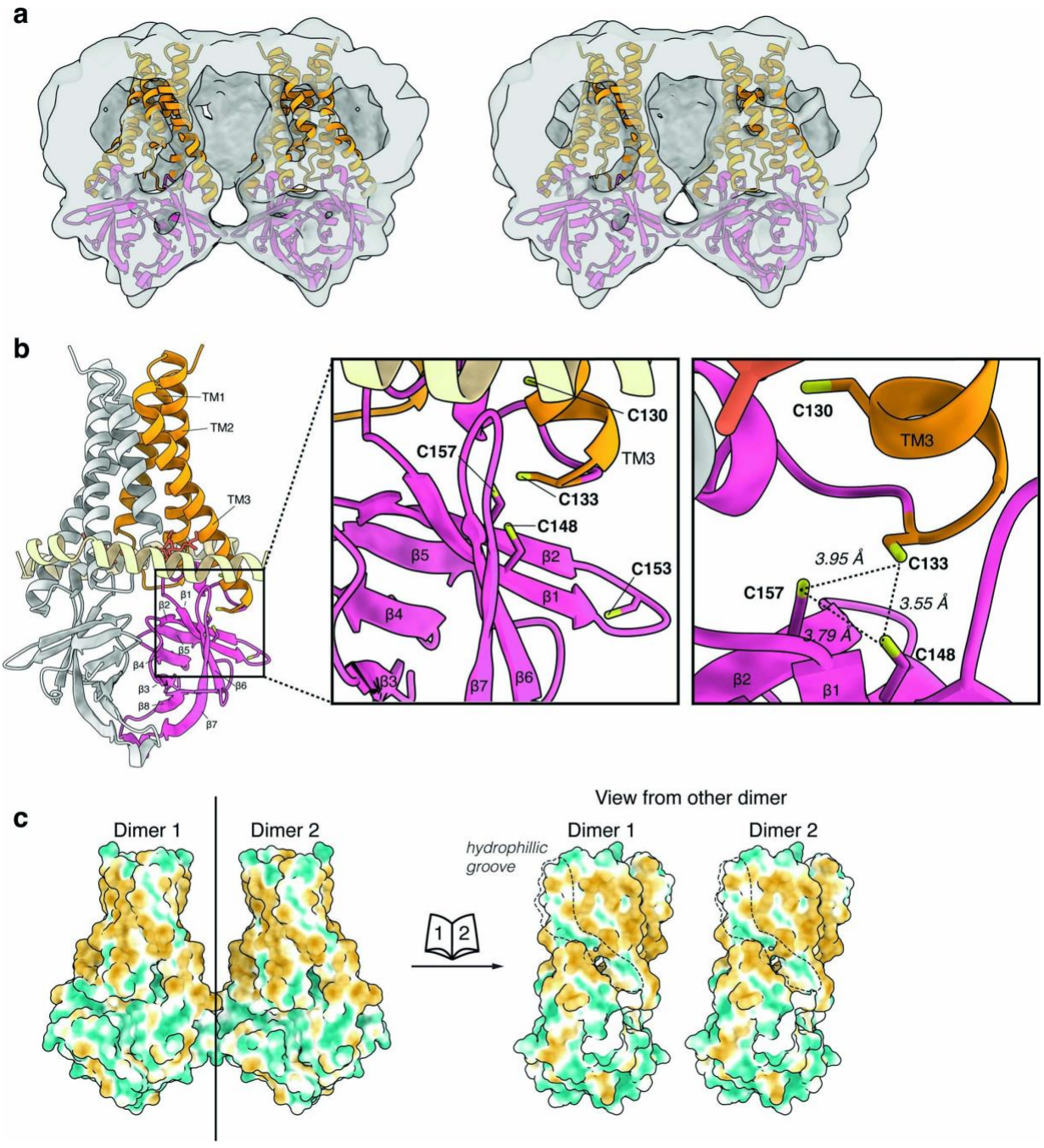


2.08 Å

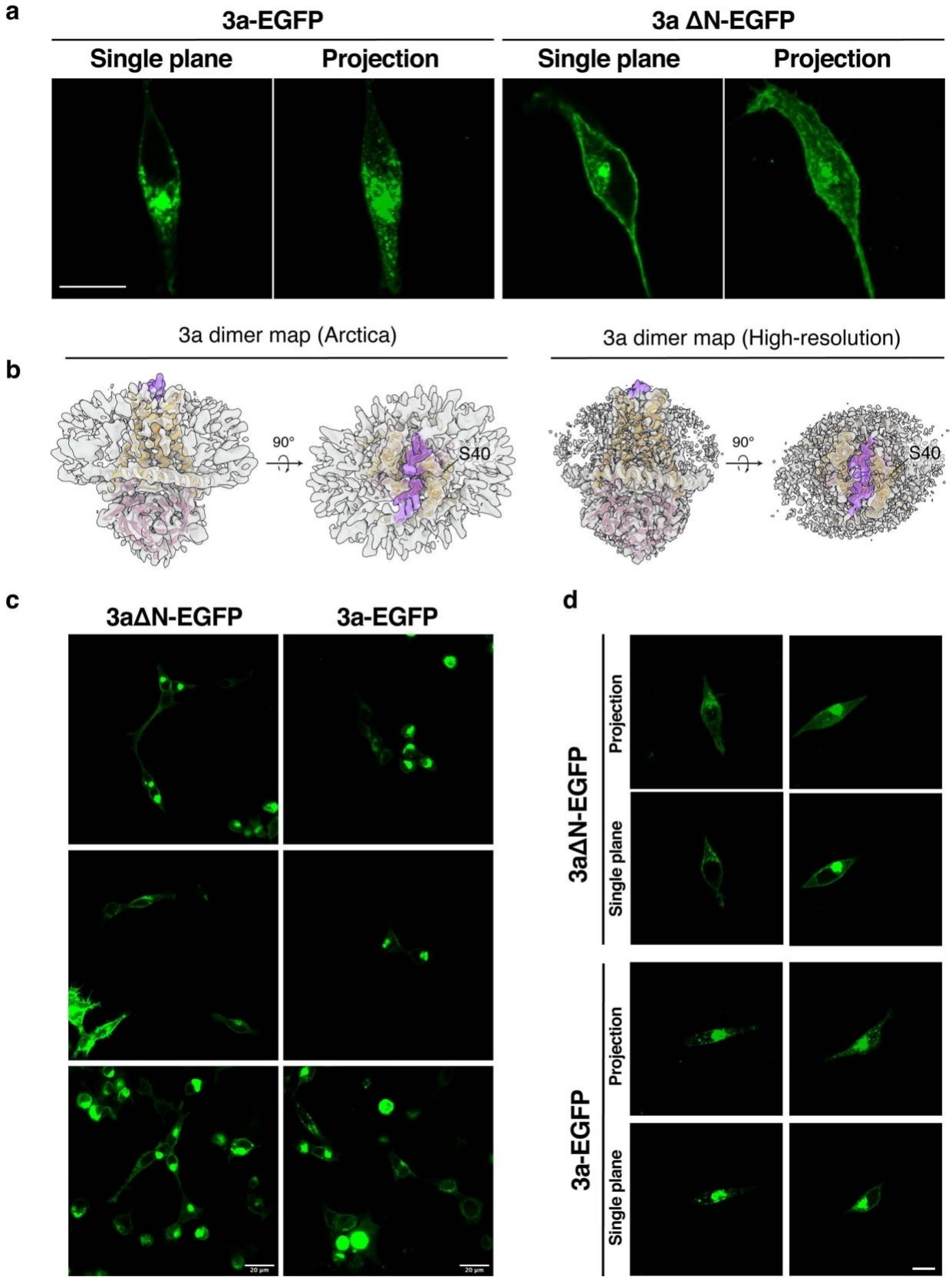
Phenix Resolve
Density Modification



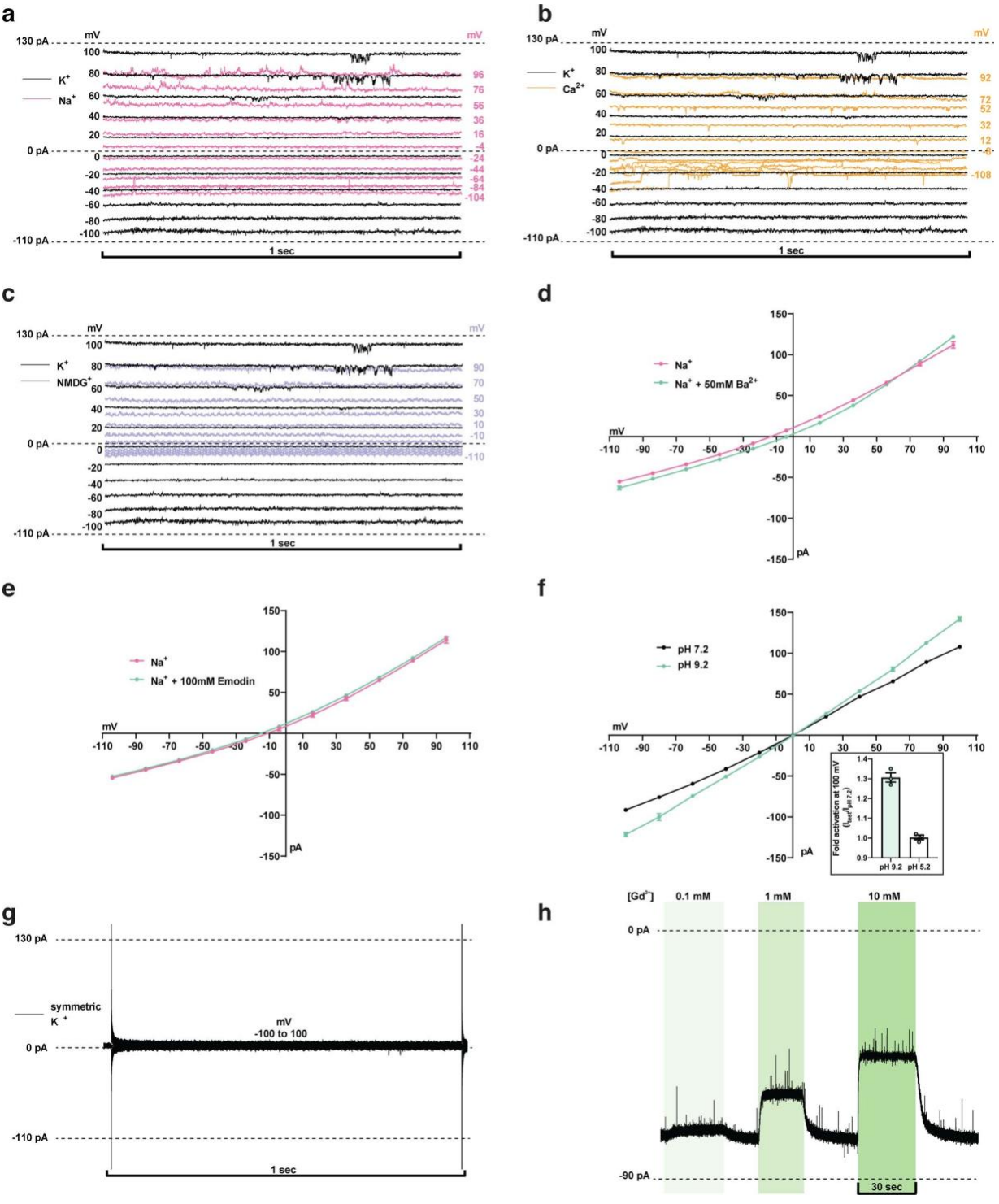
Supplemental Figure 2.9: Cryo-EM processing pipeline for 2.1 Å dimeric apo 3a in MSP1E3D1 lipid nanodiscs. Overview of Cryo-EM data processing pipeline in cryoSPARC and Relion. See Methods for details.



Supplemental Figure 2.10: The dimer-dimer interface and a cysteine rich pocket in 3a. a, A slice through the tetramer map with docked dimer models to illustrate the rigid body fit of model to map. Left: Same threshold as in Fig. 2.2 (0.25). Right: At a lower threshold (0.21). **b**, Dimer model (left) with boxed region zoom-in (middle) and an alternate view (right) to show the cysteine⁻rich region of 3a. Distances (dotted lines) between the reduced cysteines are displayed. **c**, Left: Dimers docked into the tetramer map with the solvent excluded surface of 3a colored from hydrophilic (dark cyan) to hydrophobic (dark orange). Right: 'Open book' view of each dimer from the perspective of the other dimer in the tetrameric assembly. The hydrophilic groove between TM2 and TM3 is outlined.

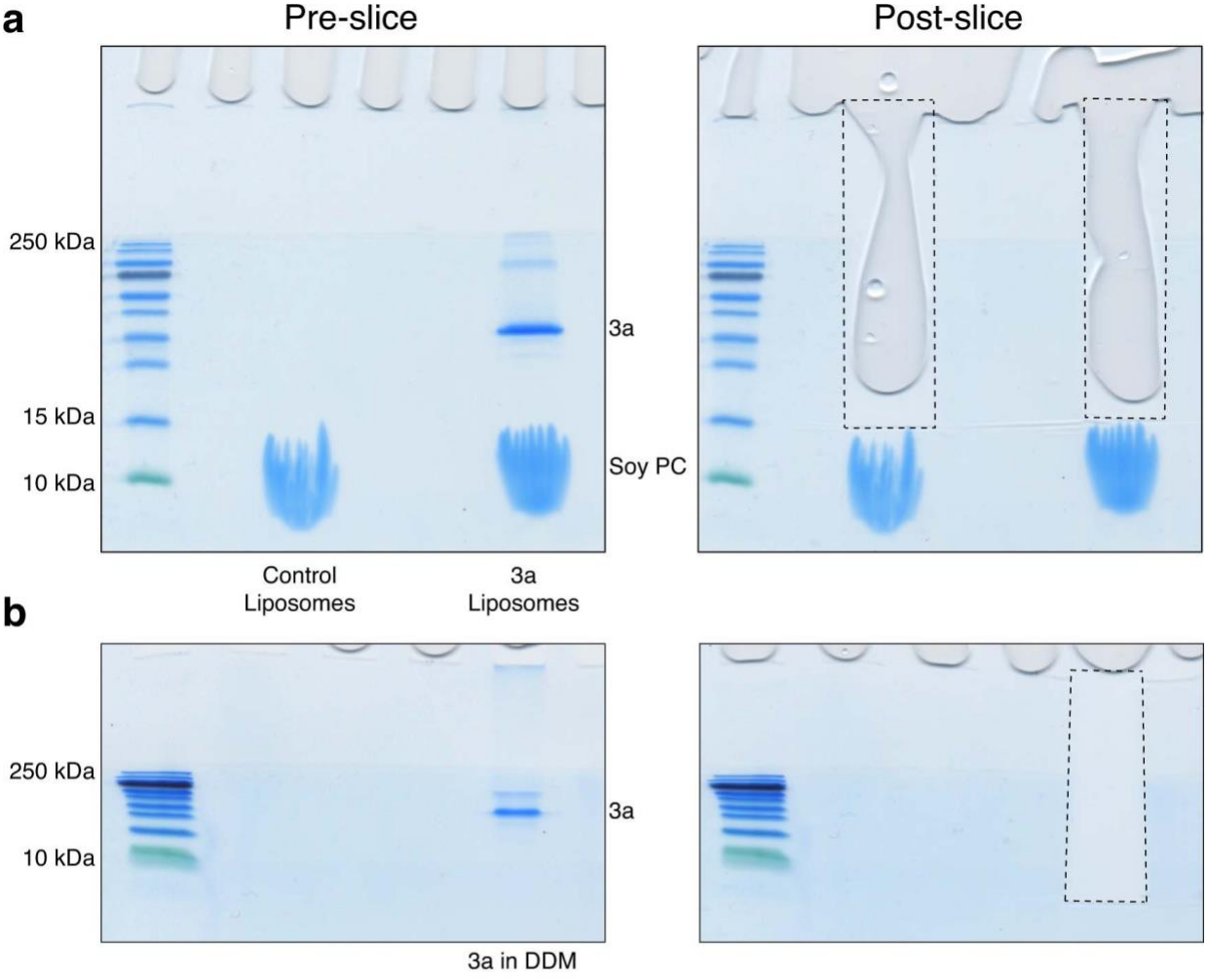

a
3a-EGFP
3a ΔN-EGFP
Single plane
Projection
Single plane
Projection
b
3a dimer map (Arctica)
3a dimer map (High-resolution)
90°
S40
90°
S40
c
3aΔN-EGFP
3a-EGFP
d
3aΔN-EGFP
Projection
Single plane
3a-EGFP
Projection
Single plane

Supplemental Figure 2.11: Localization of 3a-EGFP and 3aΔN-EGFP expressed in HEK293 cells. a, 3a-GFP fluorescence localization in HEK293 cells transfected with 3a-EGFP or 3aΔN-EGFP. Single plane and brightest-point projection are displayed for each. Scale bar, 10 μm. **b**, Side view and view from the extracellular/lumenal space for dimeric 3a cryo-EM density (gray) from the original (left) and high-resolution (right) maps with unmodeled extended density above the mouth of the pore that may correspond to the N-terminal regions colored in purple. A 3a dimer model is drawn orange (TMD) and pink (CD) inside the density. The position of the final modeled N-terminal residue (S40) is indicated. **c**, 3a-EGFP or 3aΔN-EGFP field of view with multiple cells imaged using a 20X objective. Scale bar, 20 μm. **d**, Additional images of cells imaged with the 63X objective with both single plane and brightest-point projections displayed. Scale bar, 10 μm.



Supplemental Figure 2.12: Patch recordings from 3a-proteoliposomes. (**a**-**c**,**g**) Representative current recordings from 3a-proteoliposome. Currents were recorded with the following protocol: V_{hold} = 0 mV, V_{test} = -100 to 100 mV, ΔV = 20 mV, t_{test} = 1 sec. Voltages indicated were corrected after recording for liquid junction potential. **a**, 150 mM K^+ (black) or 150 mM Na^+ (pink) bath solution. **b**, 150 mM K^+ (black) or 75 mM Ca^{2+} (orange) bath solution. **c**, K^+(black) or 150 mM $NMDG^+$ (blue) bath solution. **d**, Ba^{2+} does not block 3a currents. Current-voltage relationship plotted from a recording in 150 mM Na^+ (pink) or 150 mM Na^+ with 50 mM Ba^{2+}(green) bath solution. **e**, Emodin does not block 3a currents. Current-voltage relationship plotted from a recording in 150 mM

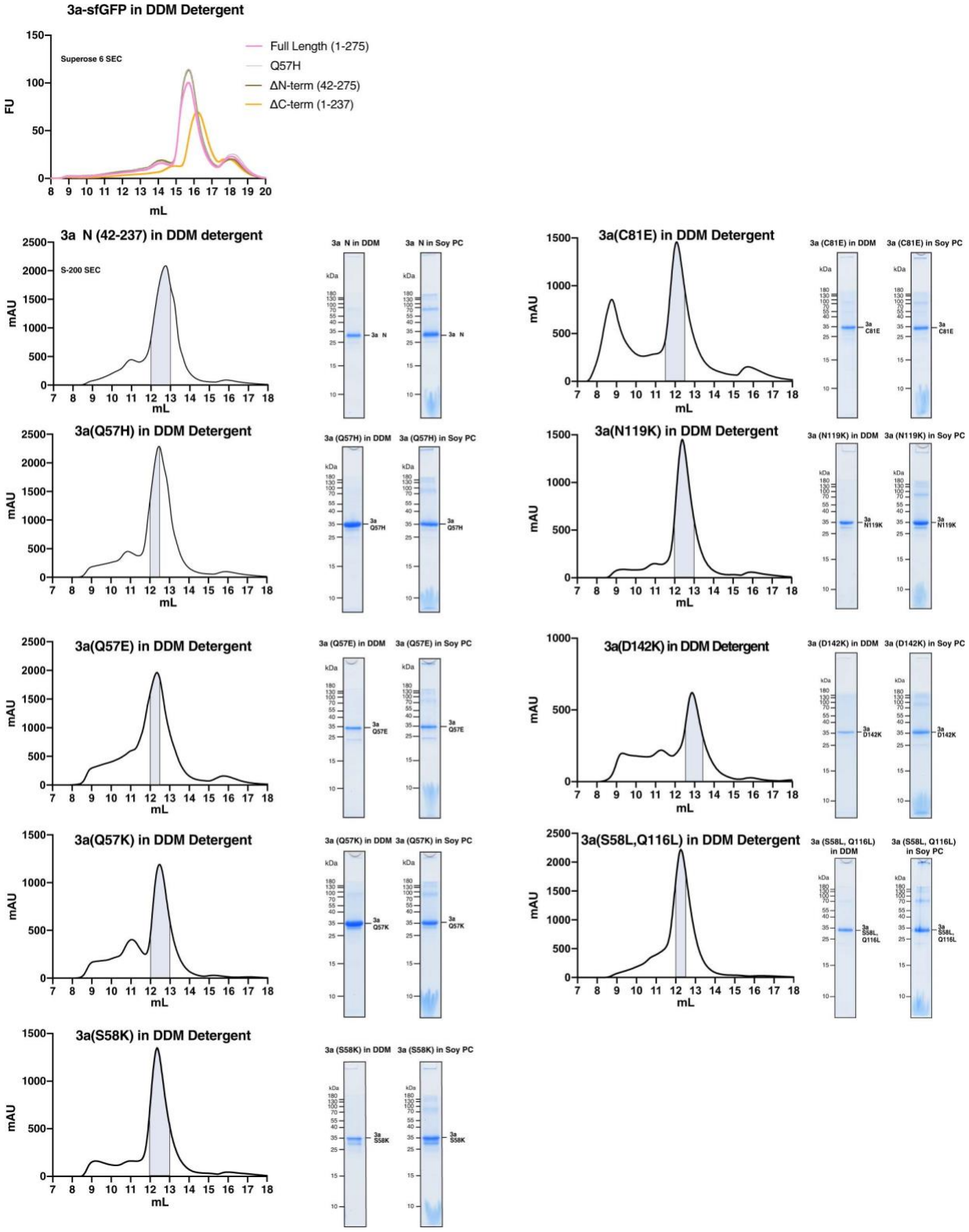
Na^+ (pink) or 150 mM Na^+ with 100 μM emodin (green) bath solution. **f**, pH sensitivity of 3a. Current-voltage relationship plotted from a recording in 150 mM K^+ pH 7.2 (black) and 150 mM K^+ pH 9.2 (green) bath solution. (inset) Fold activation at +100 mV at pH 9.2 (mean ± s.e.m., n = 3) or pH 5.2 compared to pH 7.2 (mean ± s.e.m., n = 3). **g**, Representative current recordings from mock reconstituted (empty) liposome patch. **h**, Gap-free current recording in symmetric 150 mM KCl held at −80 mV during bath solution exchange and washout of varying Gd^{3+} concentrations represented by vertical green bars. Source data for parts d, e, and f are available online.



c

Protein	Source (search)	MW (kDa)	% Sequence Coverage (# Matched Peptides)		
			3a in DDM	3a in liposomes	Control liposomes
orf3a	SARS-CoV-2 (Recombinant)	32.1	30.6% (180)	8.8% (25)	0% (0)
cytochrome P450	*S. frugiperda*	55.9	7.3% (2)	0% (0)	0% (0)
vps28	*S. frugiperda*	24.8	0% (0)	0% (0)	7.6% (1)
ubiquitin	*S. frugiperda*	8.6	52.6% (22)	0% (0)	0% (0)
ubiquitin-activating enzyme 7	Contaminant	111.7	0% (0)	0% (0)	1% (1)

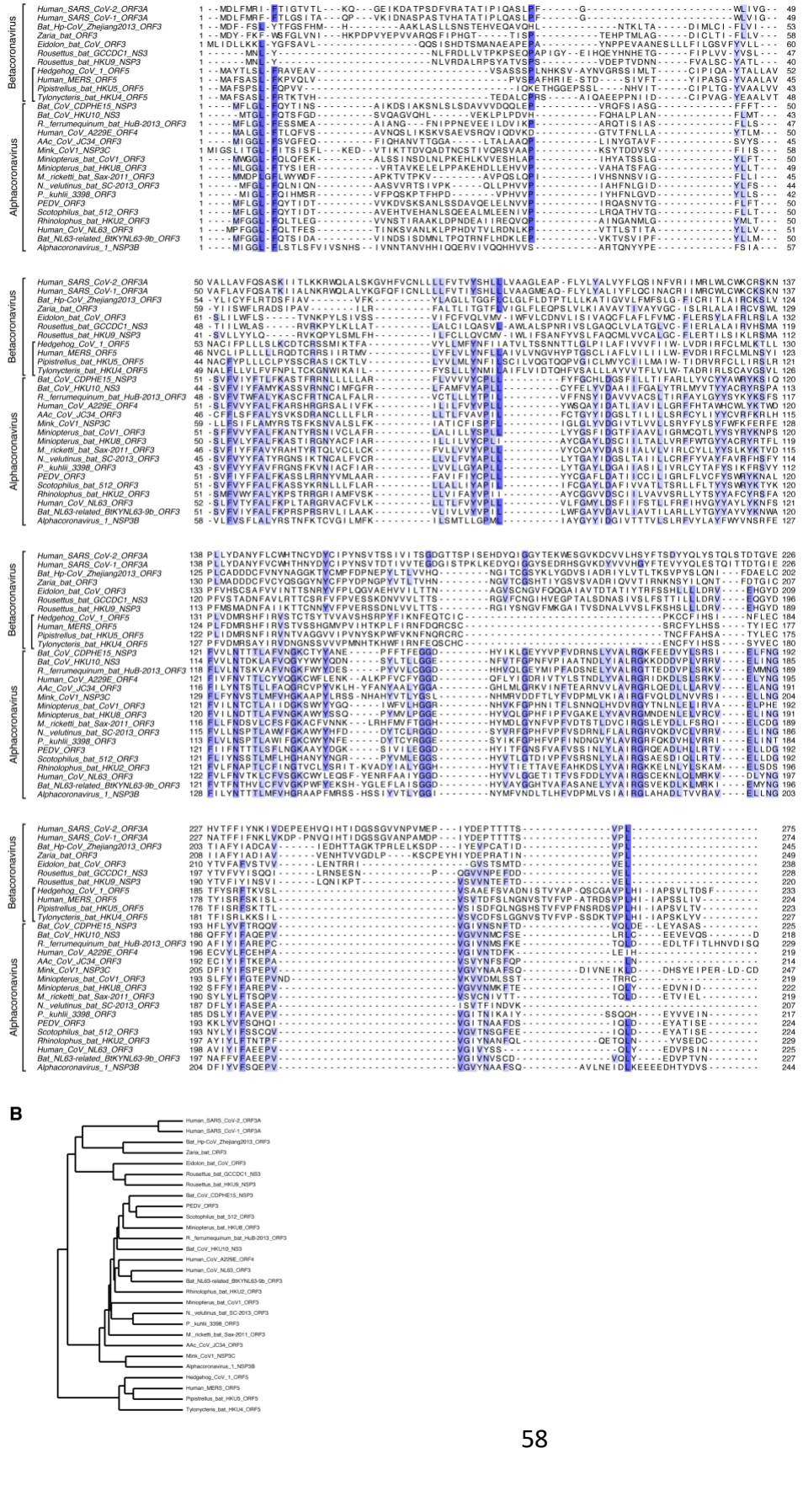
Supplemental Figure 2.13: Mass spectrometry analysis of 3a and control samples. a, SDS-PAGE gels with 3a and mock liposome samples prior to (left) and post (right) gel slice extraction for mass spectrometry. Approximate outline of the extracted region is shown with a dotted line. The control and 3a samples were from a liposome preparation used for both electrophysiological and fluorescence assays. **b**, Same as (**a**), but for 3a in DDM sample (prior to liposome incorporation). This sample was from a DDM prep used for both cryo-EM and proteoliposome preparations. **c**, Summary table of mass spectrometry analysis for the three analyzed samples. The only other hits were to human keratin proteins from the common contaminant library. No additional hits were detected from the *G. max* (soy) search.



Supplemental Figure 2.14: Purification and liposome reconstitution of 3a mutants and truncation. Top: GFP fluorescence chromatogram (FSEC) of 3a, as well as indicated mutants and truncations, expressed in SF9 cells and extracted in DDM detergent. Samples were run on a Superose 6 column. For all other panels: Size exclusion chromatogram from a s200 column of indicated 3a constructs expressed in insect cells and extracted and purified in

DDM (left), coomassie-stained SDS-PAGE of pooled dimeric 3a construct -containing fractions (center), and of 3a following reconstitution into PC lipids (right). These gels and chromatograms are from the preparations used for electrophysiological assays.

A


Betacoronavirus
Alphacoronavirus
Human_SARS_CoV-2_ORF3A
Human_SARS_CoV-1_ORF3A
Bat_Hp-CoV_Zhejiang2013_ORF3
Zaria_bat_ORF3
Eidolon_bat_CoV_ORF3
Rousettus_bat_GCCDC1_NS3
Rousettus_bat_HKU9_NSP3
Hedgehog_CoV_1_ORF5
Human_MERS_ORF5
Pipistrellus_bat_HKU5_ORF5
Tylonycteris_bat_HKU4_ORF5
Bat_CoV_CDPHE15_NSP3
Bat_CoV_HKU10_NS3
R_ferrumequinum_bat_HuB-2013_ORF3
Human_CoV_A229E_ORF4
AAc_CoV_JC34_ORF3
Mink_CoV1_NSP3C
Miniopterus_bat_CoV1_ORF3
Miniopterus_bat_HKU8_ORF3
M_ricketti_bat_Sax-2011_ORF3
N__velutinus_bat_SC-2013_ORF3
P_kuhlii_3398_ORF3
PEDV_ORF3
Scotophilus_bat_512_ORF3
Rhinolophus_bat_HKU2_ORF3
Human_CoV_NL63_ORF3
Bat_NL63-related_BtKYNL63-9b_ORF3
Alphacoronavirus_1_NSP3B

B

Human_SARS_CoV-2_ORF3A
Human_SARS_CoV-1_ORF3A
Bat_Hp-CoV_Zhejiang2013_ORF3
Zaria_bat_ORF3
Eidolon_bat_CoV_ORF3
Rousettus_bat_GCCDC1_NS3
Rousettus_bat_HKU9_NSP3
Bat_CoV_CDPHE15_NSP3
PEDV_ORF3
Scotophilus_bat_512_ORF3
Miniopterus_bat_HKU8_ORF3
R_ferrumequinum_bat_HuB-2013_ORF3
Bat_CoV_HKU10_NS3
Human_CoV_A229E_ORF4
Human_CoV_NL63_ORF3
Bat_NL63-related_BtKYNL63-9b_ORF3
Rhinolophus_bat_HKU2_ORF3
Miniopterus_bat_CoV1_ORF3
N_ velutinus_bat_SC-2013_ORF3
P_kuhlii_3398_ORF3
M_ricketti_bat_Sax-2011_ORF3
AAc_CoV_JC34_ORF3
Mink_CoV1_NSP3C
Alphacoronavirus_1_NSP3B
Hedgehog_CoV_1_ORF5
Human_MERS_ORF5
Pipistrellus_bat_HKU5_ORF5
Tylonycteris_bat_HKU4_ORF5

Supplemental Figure 2.15: Sequence alignment of 3a-like proteins across Coronaviridae a, Alignment of twenty-eight 3a-like protein sequences colored by conservation in a ramp from white (not conserved) to dark blue (highly conserved). Accession numbers are listed in Table 2. Sequences were selected from representative species from each Coronavirus subgenus. Alphacoronavirus and Betacoronavirus clades are indicated. Within Betacoronavirus the subgenus Merbecovirus is also indicated with a bar. (b) Neighbor-joining tree calculated from the alignment in (a).

Chapter 3 Structures of tweety homolog proteins TTYH2 and TTYH3 reveal a Ca^{2+}-dependent switch from intra- to intermembrane dimerization

This chapter is adapted, with permission, from work previously published: Li, B., Hoel, C.M. & Brohawn, S.G. Structures of tweety homolog proteins TTYH2 and TTYH3 reveal a Ca^{2+}-dependent switch from intra- to intermembrane dimerization. *Nat Commun* **12,** 6913 (2021). https://doi.org/10.1038/s41467-021-27283-8

3.1 Abstract

Tweety homologs (TTYHs) comprise a conserved family of transmembrane proteins found in eukaryotes with three members (TTYH1-3) in vertebrates. They are widely expressed in mammals including at high levels in the nervous system and have been implicated in cancers and other diseases including epilepsy, chronic pain, and viral infections. TTYHs have been reported to form Ca^{2+}- and cell volume-regulated anion channels structurally distinct from any characterized protein family with potential roles in cell adhesion, migration, and developmental signaling. To provide insight into TTYH family structure and function, we determined cryo-EM structures of *Mus musculus* TTYH2 and TTYH3 in lipid nanodiscs. TTYH2 and TTYH3 adopt a previously unobserved fold which includes an extended extracellular domain with a partially solvent exposed pocket that may be an interaction site for hydrophobic molecules. In the presence of Ca^{2+}, TTYH2 and TTYH3 form homomeric cis-dimers bridged by extracellularly coordinated Ca^{2+}. Strikingly, in the absence of Ca^{2+}, TTYH2 forms trans-dimers that span opposing membranes across a ~130 Å intermembrane space as well as a monomeric state. All TTYH structures lack ion conducting pathways and we do not observe TTYH2-dependent channel activity in cells. We conclude TTYHs are not pore forming subunits of anion channels and their function may involve Ca^{2+}-dependent changes in quaternary structure, interactions with hydrophobic molecules near the extracellular membrane surface, and/or association with additional protein partners.

3.2 Introduction

Tweety homologs (TTYHs) are conserved transmembrane proteins present in all eukaryotes including three members (TTYH1, TTYH2, and TTYH3) in humans[1,2,3,4,5]. The founding member, *Drosophila melanogaster* tweety (TTY), was named after a flightless cartoon bird due to its presence in the *flightless* genomic locus where mutations result in loss of flying ability[6,7], though ablation of the *tty* gene without concurrent loss of the *flightless* gene does not result in this phenotype. TTYHs have been reported to form Ca^{2+}- and cell volume-regulated anion channels and to contribute to cell adhesion, migration, and developmental signaling. In addition, TTYHs have been implicated in multiple diseases including cancers. Still, the structures, functions, and physiological roles of TTYHs have remained unclear.

Physiologically, TTYH2 and TTYH3 are broadly expressed, while expression of TTYH1 is primarily limited to the nervous system, testes, and stem cells[4,8,9,10]. Dysregulation of TTYHs expression has been implicated in some cancers: TTYH2 is upregulated in renal cell carcinoma[2], colon carcinomas[11], and osteosarcomas[12]; TTYH1 is expressed in gliomas where it contributes to brain colonization[13] and a TTYH1-C19MC microRNA cluster genetic fusion causes embryonal tumors with multilayered rosettes[10]; and TTYH3 is upregulated in gastric cancer with higher expression correlated with poor clinical outcomes[14] and a TTYH3-BRAF genetic fusion causes glioblastoma multiforme[15]. TTYHs have also been implicated in other pathologies. TTYH2 in myeloid cells interacts with SARS-CoV-2 Spike and may contribute to ACE2-independent immune cell infection and immune hyperactivation in severe cases of COVID-19[16]. TTYH1 is upregulated following epileptic events in central neurons and glia[17,18] and following inflammation in peripheral sensory neurons where it contributes to nociception and pain sensitization[19].

TTYHs are most often described as Ca^{2+}- and/or volume-regulated anion channels[5]. However, the evidence in support of channel activity to date is limited to five reports with inconsistent conclusions[3,20,21,22,23]. In the first study, TTYH2 and TTYH3 were reported to generate calcium-activated anion channels and TTYH1 was reported to generate a large conductance, calcium-independent volume-regulated anion channel[3,24]. A second study reported TTYH1 generates Ca^{2+}-dependent and volume-regulated anion channels[20]. A third study reported TTYH2 generates β-COP-regulated anion currents in low internal ionic strength and ascribed endogenous anionic currents in LoVo cancer cells to TTYH2[21]. A fourth study attributed endogenous anionic currents in SNU-601, HepG2, and LoVo cancer cells to TTYH2 and/or TTYH1 channel activity[23]. Finally, a fifth study reported TTYH1-3 form AQP4-dependent, Ca^{2+}-insensitive volume-regulated anion channels activated by cell swelling in astrocytes or in cultured cells upon cotransfection[22]. Among these studies, two reports identified the small molecule DIDS as a channel blocker[3,20] and three reported point mutations with modest effects on ion selectivity ($TTHY1_{R371Q}$, $TTYH3_{H370D}$, and $TTYH3_{R367Q}$)[3,20], DIDS block ($TTHY1_{F394S}$)[20], or current magnitude ($TTYH1_{R165A/R165C}$)[22](numbering refers to human sequences).

Other work has implicated TTYHs in aspects of cell migration, cell adhesion, and developmental signaling. TTYH2 expression in two cell lines results in increased cell aggregation[11]. TTYH1 is implicated in maintenance of neural stem cells through positive feedback between TTYH1 and Notch signaling[25,26]. Heterologous expression of TTYH1 in cultured cells induced growth of filopodia[4,17] in which TTYH1 colocalized with α5 integrin[4]. Endogenous TTYH1 was localized to neurites in rat hippocampal neurons[4,17] and to extending processes of glia[18]. In cancer cells of glial origin, TTYH1 localized to tips of tumor microtubes, neurite-like long membranous extensions involved in proliferation, invasion, and network formation among glioma cells[13]. Finally, knockdown of TTYH1 within the tumor cells resulted in altered morphology of cellular protrusions, resulting in “beading” of neurites and glioma tumor microtubes, and reduction of tumor growth and improved survival in vivo[13,17].

Intrigued by reports that TTYHs form Ca^{2+}- or volume-regulated anion channels in the absence of homology to known ion channels, we set out to characterize TTYH structure and function. Here, we present cryo-EM structures of TTYH2 and TTYH3 and a functional analysis of TTYH2. We conclude TTYHs are not pore-forming subunits of ion channels, but adopt a previously unobserved fold capable of Ca^{2+}-dependent changes in oligomeric state and quaternary structure.

3.3 Results

3.3.1 Structure determination

Full-length *Mus musculus* TTYH2 was expressed in HEK293T cells with a cleavable C-terminal fusion to EGFP, extracted and purified in detergent, and reconstituted into lipid nanodiscs formed by the scaffold protein MSP1E3D1 and lipids in the presence of 1 mM Ca^{2+} (Supplemental Figure 3.1). We determined the cryo-EM structure of TTYH2 in the presence of Ca^{2+} to 3.3 Å resolution (Fig. 3.1 and Supplemental Figure 3.2). The map enabled de novo modeling of 396 of 532 amino acids from each 59 kDa TTYH2 protomer chain with the majority of unmodeled residues within the unstructured C-terminal region (amino acids 415–532) and

others within poorly resolved segments of the N-terminus and intracellular loops. TTYH2 adopts a previously unobserved fold among proteins of known structure; Dali[27] searches using intact TTYH2 or isolated transmembrane or extracellular regions against the protein structure database do not return high confidence structural homologs.

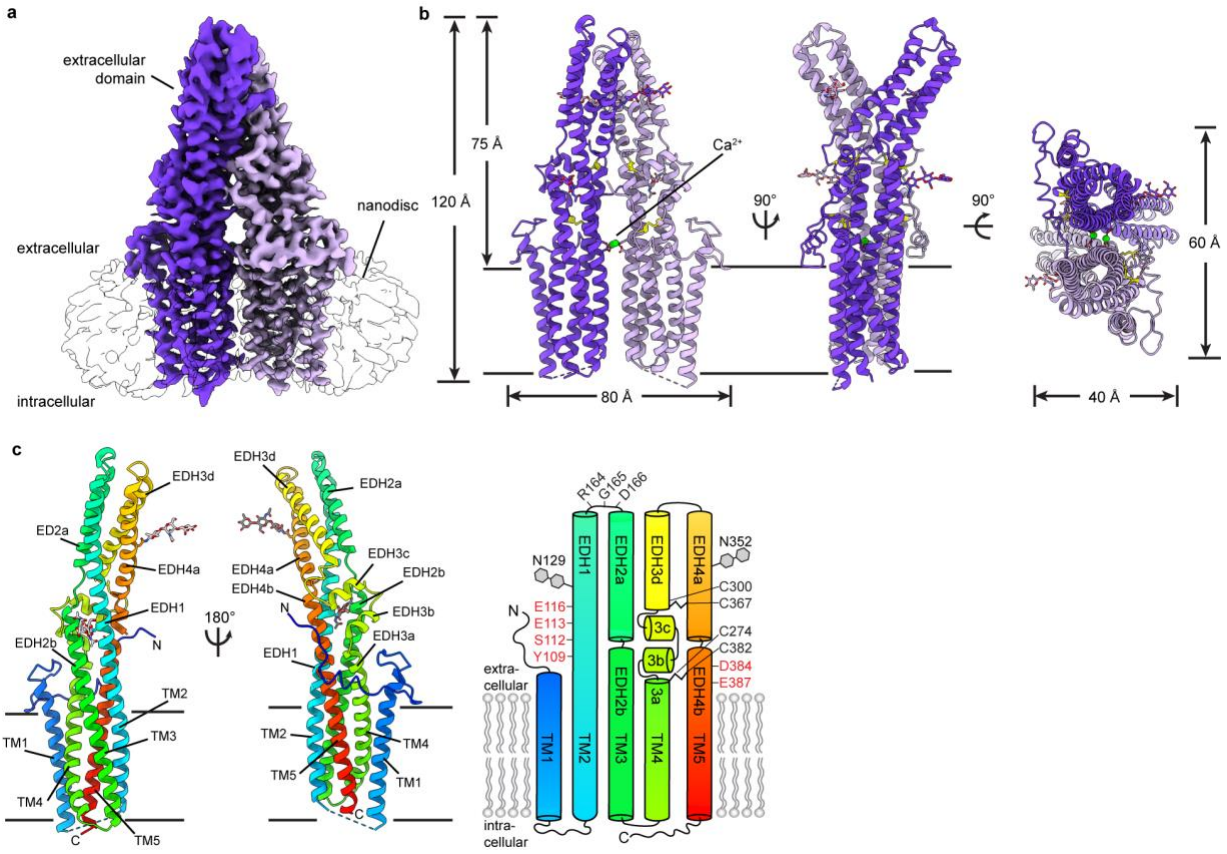


Figure 3.1: Structure of a TTYH2 *cis*-dimer in the presence of Ca^{2+}. a Cryo-EM map of a TTYH2 *cis*-dimer in MSP1E3D1 nanodiscs in the presence of Ca^{2+} at 3.3 Å resolution viewed from the membrane plane. Density from the TTYH2 protomers is colored dark and light purple and the nanodisc is white. **b** Model of the TTYH2 *cis*-dimer viewed from the membrane and extracellular side. N-linked glycosylation sites and disulfide bonds in the extracellular domain are drawn as sticks and Ca^{2+} ions are shown as green spheres. **c** A TTYH2 protomer viewed from the membrane plane and cartoon illustration of domain topology with rainbow coloring from blue (N-terminus) to red (C-terminus). Positions of disulfide bonds, glycosylation sites, and residues discussed in the text are indicated.

3.3.2 TTYH2 and TTYH3 *Cis*-dimer architecture

In the presence of calcium, TTYH2 adopts a side-by-side homodimer (a "*cis*-dimer") within a single nanodisc membrane (Fig. 3.1). Each protomer consists of five transmembrane helices (TM1–5) and an extracellular domain (ED) which extends 75 Å from the membrane surface between an extracellular N-terminus and intracellular C-terminus. The ED largely consists of a kinked bundle of four helices (EDH1–4). Two N-linked glycosylation sites are observed and partially modeled at N129 on EDH1 and N352 on EDH4. EDH1 is a relatively unbroken helix extending from TM2, EDH2 and EDH4 are split into two segments each (a and b) by helical breaks, and EDH3 is broken into four segments (a–d) with EDH3b and EDH3c forming a short helix-turn-helix motif positioned with their helical axes approximately perpendicular to those of

EDH3a and EDH3d. The ED appears rigidly packed and is stabilized above and below EDH3b and EDH3c by disulfide bonds between EDH3d and EDH4a (between C300 and C367) and EDH3a and EDH4b (between C274 and C382). The TTYH2 protomer architecture is in close accordance with one investigation of TTYH topology[28] and in contrast to other predictions[3,22,24].

The high degree of homology between TTYHs (*M. musculus* TTYH1-3 display ~40% identity across their sequence), suggests a common molecular architecture. To test this, we investigated the structure of *M. musculus* TTYH3. TTYH3 was expressed, purified, and reconstituted into lipid nanodiscs in the presence of 1 mM Ca^{2+} similarly to TTYH2 (Supplemental Figure 3.1). A cryo-EM reconstruction of TTYH3 was determined to 3.2 Å resolution and the de novo modeled structure was compared to TTYH2 (Supplemental Figure 3.3). The overall protomer architecture and *cis*-dimerization arrangement was nearly identical between TTYH2 and TTYH3 (overall r.m.s.d. = 1.4 Å). The most substantial difference is the absence of TM1 from one protomer in TTYH3: we found during data processing that TTYH3 reconstructions were clearly asymmetric with thinner nanodisc density on one side of the complex. This thinner disc density displayed weak or absent density for TM1. While a subset of particles in the TTYH2 dataset were similarly asymmetric (Supplemental Figure 3.3d), most displayed symmetric and interpretable density for TM1s in each protomer. Whether this difference is functionally relevant remains to be determined. We focus our discussion on the TTYH2 structure unless otherwise noted.

Association between TTYH2 protomers in the *cis*-dimer buries an interface of 1556 $Å^2$ and involves residues in the transmembrane and EDs, primarily in TM2, TM5, EDH1, EDH3d, EDH4a, and EDH4b. A striking feature of the interface is the close juxtaposition of a surface formed by conserved negatively charged or electronegative residues from each protomer in the ED just above the membrane surface (Fig. 3.2a-c). This region includes residues from EDH1 (Y109, S112, E113, and E116) and EDH4b (D384 and E387) with Y109 and S112 underneath E113, E116, D384, and E387. The cryo-EM map contains strong density in the space between the charged residues bridging opposing protomers (Fig. 3.2a). Based on electrostatic and geometric features of this site, we speculate that these densities correspond to coordinated Ca^{2+} ions that would reduce repulsion otherwise expected to occur between the electronegative surfaces of each protomer.

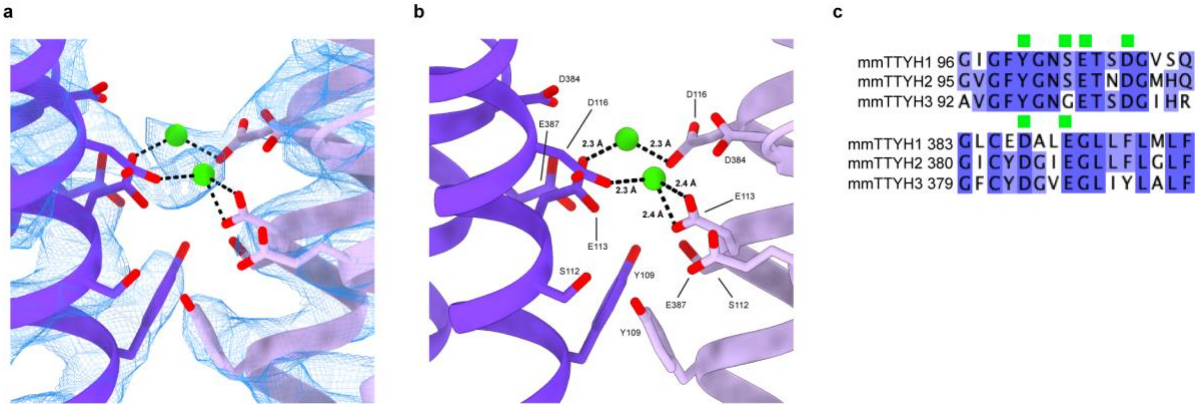


Figure 3.2: Extracellular $Ca2^+$ binding sites in TTYH2. a, **b** View from the membrane plane of the TTYH2 *cis*-dimer model (purples) and cryo-EM density (blue mesh). Electronegative side chains from each protomer are displayed as sticks and Ca^{2+} ions are shown as green spheres. **c** Alignment of mouse TTYHs with electronegative

residues near the Ca^{2+} binding sites indicated with green boxes and conservation colored from white (not conserved) to blue (most conserved).

3.3.3 Ca^{2+}, monomer, and *trans*-dimer

We reasoned that if Ca^{2+} was an important component of the *cis*-dimer interface for bridging the electronegative surfaces of each protomer, removing Ca^{2+} from TTYH2 could alter its oligomeric state. To test this, we purified and reconstituted TTYH2 into lipid nanodiscs without Ca^{2+} and in the presence of EGTA to maintain low free Ca^{2+} concentrations for structure determination. We also pretreated filter paper used for EM grid blotting with EGTA to sequester the high residual Ca^{2+} known to be present in this material to prevent Ca^{2+} transfer to the TTYH2 nanodisc sample[29].

The Ca^{2+}-free TTYH2 cryo-EM data did reveal a change in oligomeric state, but surprisingly we identified two different conformations in approximately equal proportions: the expected monomeric TTYH2 (to 4.0 Å resolution) and an unexpected head-to-head homodimer (a "*trans*-dimer" to 3.9 Å resolution) in which TTYH2 protomers embedded in separate nanodiscs associate via a buried surface at the distal end of each ED (Fig. 3.3 and Supplemental Figure 3.4). The *trans*-dimer therefore bridges two opposing membranes over a ~130 Å extracellular space. Protomers within monomeric, *trans*-dimeric, and *cis*-dimeric TTYH2 adopt overall very similar conformations (pairwise protomer r.m.s.d. = 1.3–1.5 Å) with conformational changes largely limited to subtle rearrangements of side chains (Supplemental Figure. 3.5). The *trans*-dimeric interface is smaller than the *cis*-dimeric interface, burying 908 $Å^2$ compared to 1556 $Å^2$ and is largely polar with eight intersubunit hydrogen bonds (between residues D166-Q316, D166-T320, Q169-T321, and Q325-Q325), π-π stacking interactions (between F173-F173), and nonpolar interactions (between I324-L329). The *trans*- and *cis*-dimeric interfaces are partially overlapping with residues Q316, R317, T320, and T321 each contributing to both TTYH2 *cis*-dimer and *trans*-dimer interactions (Fig. 3.4). As a consequence of this partial overlap in interfaces, *cis*-dimerization and *trans*-dimerization are expected to be mutually exclusive. Notably, residues involved both *cis*- and *trans*-dimerization interfaces are highly conserved with markedly lower conservation evident on the opposing face of the molecule that is not involved in oligomerization (Fig. 3.4).

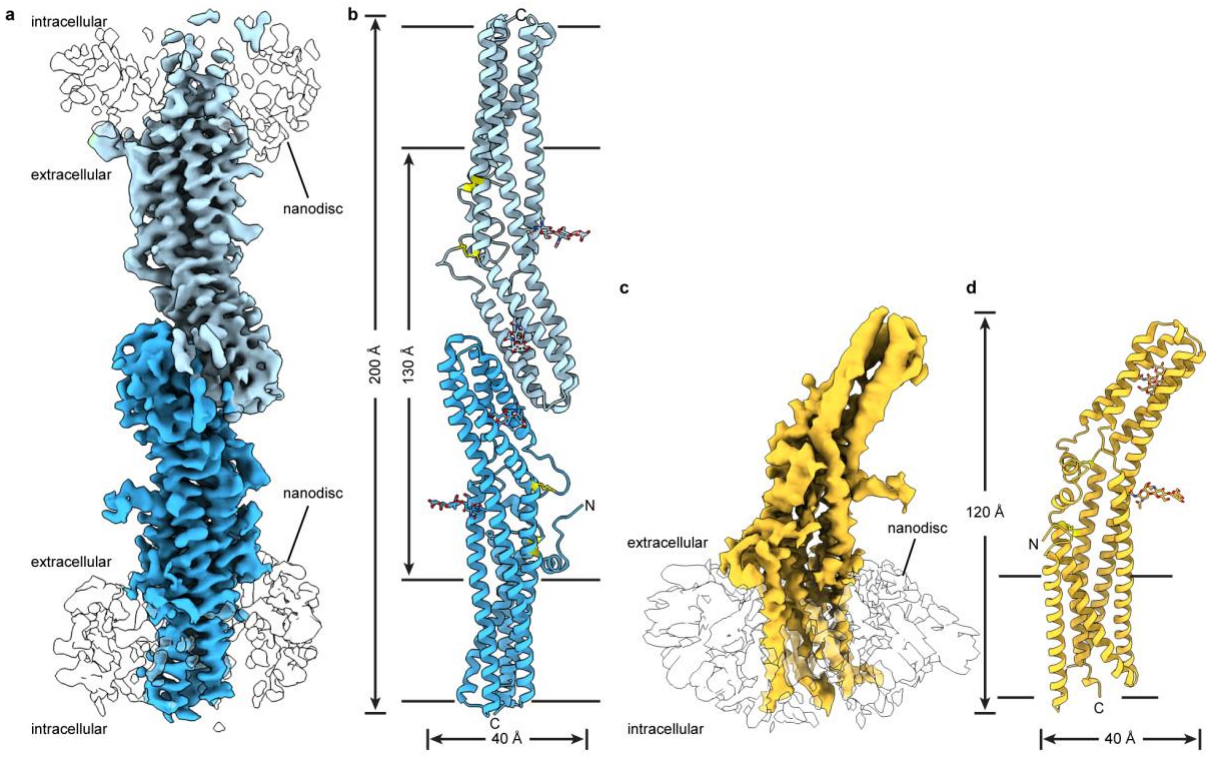


Figure 3.3: Structures of a TTYH2 *trans*-dimer and monomer in the absence of Ca^{2+}. a Cryo-EM map of a TTYH2 *trans*-dimer in MSP1E3D1 nanodiscs in the absence of Ca^{2+} at 3.9 Å resolution viewed from the membrane plane. Density from the TTYH2 protomers is colored dark and light blue and the nanodiscs are white. **b** Model of the TTYH2 *trans*-dimer viewed from the membrane. N-linked glycosylation sites and disulfide bonds in the extracellular domain are drawn as sticks. **c** Cryo-EM map of a TTYH2 monomer in MSP1E3D1 nanodiscs in the absence of Ca^{2+} at 4.0 Å resolution viewed from the membrane plane. Density from the TTYH2 protomer is colored orange and the nanodisc is white. **d** Model of the TTYH2 monomer viewed from the membrane. N-linked glycosylation sites and disulfide bonds in the extracellular domain are drawn as sticks.

3.3.4 Structural and functional analysis

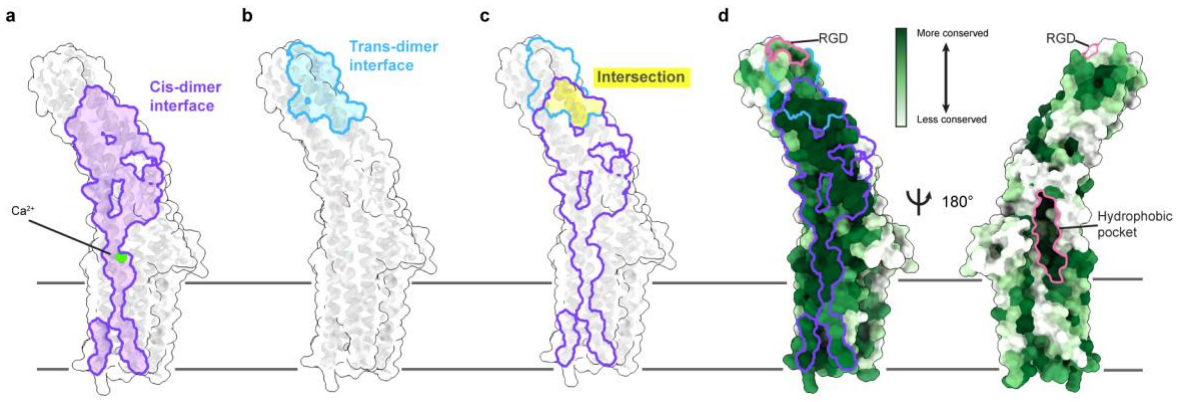


Figure 3.4: Conservation and partial overlap of the TTYH2 *cis*- and *trans*-dimerization interfaces. a "Open book" view of the molecular surface of one TTYH2 protomer. The *cis*-dimerization interface is outlined and colored purple. The position of the bound Ca^{2+} is shown. **b** Open book view of the molecular surface of one protomer from the TTYH2 *trans*-dimer. The *trans*-dimerization interface is outlined and colored blue. **c** Overlay of *cis*-dimerization

(purple outline) and *trans*-dimerization (blue outline) interfaces with intersection highlighted in yellow. **d** TTYH2 surface colored according to conservation among chordate TTYH1-3 sequences. *cis*-dimer and *trans*-dimer interfaces are outlined as in **a**–**c**. The extracellular hydrophobic pocket and RGD motif are outlined in pink.

A striking aspect of the TTYH structures is that they display no obvious path for ion conduction across the membrane as would be expected for an anion channel. One possible location for an ion conducting pore is within each protomer, as observed in other anion channels including CLCs[30]. In both TTYH2 and TTYH3, TM2, TM3, TM4, and TM5 pack tightly through a predominantly hydrophobic interface leaving no path for conduction in the absence of dramatic conformational changes. A second possible location for a pore is between subunits of a *cis*-dimer. In TTYH2 and TTYH3, this interface is less tightly packed, but almost exclusively hydrophobic, which would create a high energy barrier for ion passage. A third possible route for ion conduction is along hydrophilic grooves in the transmembrane region facing and partially exposed to the lipid bilayer, similar to those in TMEM16 family anion channels[31,32]. TTYH2 and TTYH3 display no such hydrophilic groove and rather show a hydrophobic surface across the membrane typical of most membrane proteins. In addition to the lack of a clear pore in the present structures, mapping point mutations previously reported to alter channel properties[3,20,22] onto the structures shows that these residues are unlikely to contribute to putative conduction paths: two are near the top of the EDH4b and contribute or are close to the *cis*-dimer interface (R368 and H371), one is membrane facing at the extracellular end of TM5 (F391), and one is present on the distal tip of the ED in the short linker between EDH1 and EDH2a (R164) (Supplemental Figure 3.6).

To further investigate potential ion channel activity in TTYH2, we performed electrophysiological recordings from TTYH2-expressing cells (Supplemental Figure 3.6). Consistent with the structural analysis, we were unable to observe the previously described TTYH2-dependent currents[3,20,21,22,23]. Cells expressing TTYH2 and AQP4 did not display currents significantly different from control cells under basal conditions or after perfusion of hypotonic extracellular solution that resulted in cell swelling (Supplemental Figure 3.6). Based on structural data showing the lack of putative conduction paths in TTYH2 and TTYH3 and these electrophysiological results, we conclude TTYH proteins are not pore-forming subunits of anion channels.

3.4 Discussion

TTYHs are most often described as anion channels regulated by Ca^{2+} or cell swelling[5]. However, there are only limited experimental data that support this assignment[3,20,21,22,23], some results from these studies are contradictory, and other evidence has been presented[33,34] to refute the conclusion reached in the first of these reports that TTYH1 generates maxi-anion channels upon heterologous expression in cultured cells[3]. Evidence presented here is inconsistent with TTYHs forming pore-forming subunits of anion channels. There are no apparent pathways for ion transport across the membrane within a TTYH protomer, along the outside of protomers through hydrophilic grooves, or between protomers in *cis*-dimers and we are unable to observe TTYH2 currents in cells. Although it is possible that stimuli other than Ca^{2+} or hypotonicity-induced cell swelling opens pores in TTYHs, the structural evidence leads us to hypothesize TTYHs perform other cellular functions. It may be that prior reports of TTYH-dependent channel activity[3,20,21,22,23] are a consequence of indirect effects of TTYH expression or manipulation on

other ion channels expressed in cells including volume-regulated anion channels formed by LRRC8A-E. It also remains a possibility that TTYHs serve as auxiliary subunits of other ion channels. Either of these scenarios could plausibly explain prior results including the observation that volume-regulated anion channel currents are completely ablated in TTYH1 and/or TTYH2 knockout or knockdown cells and rescued by subsequent reintroduction of TTYH1 or TTYH2[22,23].

How might we explain the physiological and pathological role of TTYHs if not through the direct or indirect conduction of anions? A notable structural feature in the ED of TTYHs suggests that interactions with lipids or other hydrophobic ligands could play a functional role. Within each TTYH protomer, we observe a large, conserved, and primarily hydrophobic pocket facing away from the *cis*-dimer interface (Fig. 3.4d and Supplemental Figure 3.7). The pocket is positioned just above the outer membrane leaflet and is open both to the bilayer and extracellular solution. In TTYH2 *cis*-dimers, we observe horseshoe-shaped density within the pocket, although the density is not sufficiently featured at this resolution to unambiguously identify and model (Supplemental Figure 3.7). The pocket and unassigned density is similarly observed in TTYH3 *cis*-dimers, TTYH2 monomers, and TTYH2 *trans*-dimers, suggesting it may be a conserved site for binding small molecules or extended and primarily apolar lipid headgroups. It is conceivable that TTYHs utilize this pocket for binding or transporting lipids, other ligands, or short peptides from interacting proteins.

TTYHs including TTYH2 have also been implicated in cell development, growth, and adhesion. Could these roles be related to functions other than anion conduction? Shortly after its discovery, it was recognized that TTYH2 contains an RGD sequence motif[2,35]. Similar motifs in cell matrix proteins mediate direct interactions with RGD-binding integrins and RGD-containing peptides and other RGD mimetics are sufficient for integrin binding[36,37]. Consistent with this notion, TTYH1 was reported to colocalize by immunofluorescence with α5 integrin in the filopodia of transfected cells[4]. In mouse TTYH2, the RGD motif (R164, G165, D166) is positioned at the distal tip of the ED within the short linker between helices EDH1 and EDH2 (Fig. 3.4d and Supplemental Figure 3.8). The RGD sequence is well conserved among TTYH2s. TTYH1 and TTYH3 sequences generally maintain positive and negative charges at positions corresponding to R164 and D166 and display more variation in the identity of the central residue. Intriguingly, one of the point mutations previously reported to impact TTYH2 ion channel activity and proposed to form part of the ion conduction pore was of the arginine in this sequence (R164)[22]. Given that this residue is ~75 Å from the membrane surface, it may more likely be involved in protein-protein interactions (Fig. 3.4d and Supplemental Figure 3.8). Indeed, the RGD motif forms part of the *trans*-dimerization interface in TTYH2. In contrast, it is exposed in TTYH2 *cis*-dimers and monomers with ~250 $Å^2$ of solvent accessible area presented to solution by each RGD. This difference suggests that if RGD or RGD-like motifs in TTYHs do interact with integrins, the interaction could be regulated by extracellular Ca^{2+} levels: *cis*-dimers favored in the presence of Ca^{2+} would display RGDs available for interaction, while *trans*-dimerization in low Ca^{2+} conditions would disfavor integrin interaction by sequestration of the RGD surface. It may therefore be that the reduction in current previously observed as a result of mutating R164[22] is a consequence of disrupting TTYH *trans*-dimerization or TTYH-integrin interactions that in turn impact cell swelling-activated channel activity.

During preparation of this manuscript, an independent report of TTYH structures (*H. sapiens* TTYH1-3) was published[38]. In that report, human TTYH1, TTYH2, and TTYH3 structures are captured as *cis*-dimers similar to mouse TTYH2 and TTYH3 *cis*-dimers reported here (Supplemental Figure 3.5, pairwise protomer r.m.s.d. = 1.2 Å). Structural features including the Ca^{2+} interaction site bridging protomers in the *cis*-dimer and the presence of a hydrophobic pocket in the ED with unmodeled density features are similarly conserved. Notably, reported human TTYH1 and TTYH2 structures determined in Ca^{2+}-free conditions are also *cis*-dimers, while we observe monomers and *trans*-dimers of mouse TTYH2 prepared in Ca^{2+}-free conditions here. It may be that human and mouse TTYH2 have a different propensity for Ca^{2+}-dependent changes in oligomerization state. However, human and mouse TTYH2 display a high degree of overall sequence (83% identity) and structural conservation, including at the *trans*-dimerization interface where 18/19 residues involved are identical between human and mouse. The only difference (mouse G327 and human A327) is not expected to impact the propensity of TTYH2 to *trans*-dimerize. We therefore do not expect sequence differences between species at the interface to account for the difference in observed oligomeric state. Alternatively, sufficient Ca^{2+} may have been present in the reported Ca^{2+}-free human TTYH1 and TTYH2 samples (perhaps due to leaching from filter paper) to occupy Ca^{2+} sites and promote *cis*-dimerization[29].

We observed Ca^{2+}-dependent changes in TTYH2 oligomerization state in structures determined at 1 mM Ca^{2+} and in the absence of Ca^{2+}. Whether these oligomeric state changes are physiologically relevant depends on several currently unknown factors. For dissociation of *cis*-dimers into monomers, these include: (1) the affinity of TTYHs for Ca^{2+}, (2) the related dependence of TTYH oligomeric state changes on $[Ca^{2+}]_{ext}$, and (3) the range of $[Ca^{2+}]_{ext}$ experienced by TTYH-expressing cells under different conditions. For association of monomers into *trans*-dimers, these additionally include: (4) the binding constant for *trans*-dimerization and (5) the surface density of natively expressed TTYHs. Future studies into these factors and potential interactions between TTYHs and integrins or other extracellular binding partners (and their Ca^{2+}-dependence) will be important for interpreting the physiological and functional roles of the TTYH structures reported here.

3.5 Methods

3.5.1 Cloning, expression, and purification

Genes encoding *M. musculus* TTYH2 and TTYH3 were codon-optimized for expression in *H. sapiens* embryonic kidney cells (HEK293S GNTI$^-$ cells) and *S. frugiperda* cells (Sf9 cells), respectively, and synthesized (Integrated DNA Technologies) (Table 3.2). The TTYH2 gene was cloned into a custom vector based on the pEG Bacmam[39] backbone with an added C-terminal PreScission protease (PPX) cleavage site, linker sequence, superfolder GFP (sfGFP) and 10×His tag, generating a construct for expression of TTYH2-SNS-LEVLFQGP-TAAAA-sfGFP-GGG-10×His. The construct was transformed into the DH10Bac *E. coli* strain to generate the recombinant bacmid DNA which was then used to transfect insect cells to generate BacMam P1 virus. P2 virus was then generated by infecting cells at two million cells per mL with P1 virus at a multiplicity of infection of roughly 0.1, with infection monitored by fluorescence and harvested after 72 h. P3 virus was generated in a similar manner. P3 viral stock was then used to infect suspension HEK293S GNTI- cells at two million cells per mL grown at 37 °C with 5% CO_2.

After 24 h, 10 mM sodium butyrate was added and cells were grown for another 48 h at 30 °C in the presence of 5% CO_2. Cells were pelleted, flash-frozen in liquid nitrogen, and stored at −80 °C.

A cell pellet from 2 L culture was thawed on ice and lysed in 100 mL of hypotonic buffer containing 10 mM Tris pH 8.0, 1 mM $CaCl_2$, 1 mM phenylmethysulfonyl fluoride, 1 mM E64, 1 mg/mL pepstatin A, 10 mg/mL soy trypsin inhibitor, 1 mM benzimidine, 1 mg/mL aprotinin, and 1 mg/mL leupeptin. The membrane fraction was collected by centrifugation at 150,000 × *g* for 45 min and homogenized with a Dounce homogenizer in 100 mL lysis buffer containing 20 mM Tris pH 8.0, 150 mM NaCl, 1 mM $CaCl_2$, 1 mM phenylmethysulfonyl fluoride, 1 mM E64, 1 mg/mL pepstatin A, 10 mg/mL soy trypsin inhibitor, 1 mM benzimidine, 1 mg/mL aprotinin, 1 mg/mL leupeptin, 10 μL Benzonase® endonuclease (EMD Millipore), 1% n-dodecyl-b-D-maltopyranoside (DDM, Anatrace, Maumee, OH), 0.2% cholesteryl hemisuccinate tris salt (CHS, Anatrace). Protein was extracted with gentle stirring for 2 h at 4 °C. Five milliliters of Sepharose resin coupled to anti-GFP nanobody was added to the supernatant and stirred gently for 2 h at 4 °C. The resin was collected in a column and washed with 50 mL buffer 1 (20 mM Tris, 500 mM NaCl, 1 mM $CaCl_2$, 0.025% DDM, 0.005% CHS, pH 8.0), 150 mL buffer 2 (20 mM Tris, 150 mM NaCl, 1 mM $CaCl_2$, 0.025% DDM, 0.005% CHS, pH 8.0), and 20 mL of buffer 1. PPX (~0.5 mg) was added into the washed resin in 5 mL buffer 2 and rocked gently overnight. Cleaved TTYH2 was eluted and concentrated to ~0.5 mL with an Amicon Ultra spin concentrator (50 kDa cutoff, MilliporeSigma, USA). The concentrated protein was subjected to size exclusion chromatography using a Superose 6 Increase 10/300 column (GE Healthcare, Chicago, IL) run in buffer 3 (20 mM Tris pH 8.0, 150 mM NaCl, 1 mM $CaCl_2$, 0.025% DDM, 0.005% CHS) on a NGC system (Bio-Rad, Hercules, CA). The peak fractions were collected, and spin concentrated for nanodisc reconstitution. For cryo-EM samples in absence of Ca^{2+}, 5 mM EGTA was added into the buffer instead of 1 mM $CaCl_2$.

The TTYH3 gene was cloned into a custom vector based on the pACEBAC1 backbone (MultiBac, Geneva Biotech) with an added C-terminal PreScission protease (PPX) cleavage site, linker sequence, superfolder GFP (sfGFP) and 7×His tag, generating a construct for expression of TTYH3-SNS-LEVLFQGP-SRGGSGAAAGSGSGS-sfGFP-GSS-7×His in Sf9 cells. The construct was transformed into the DH10MultiBac *E. coli* strain to generate recombinant bacmid DNA which was then used to transfect Sf9 to generate P1 virus. P2 virus was generated by infecting Sf9 cells at two million cells per mL with P1 and harvested after 72 h. P3 virus was generated in a similar manner. P3 viral stock was used to infect suspension Sf9 cells at two million cells per mL grown at 27 °C. After 72 h, cells were pelleted, flash-frozen in liquid nitrogen, and stored at −80 °C. A cell pellet from 1 L culture was thawed on ice and cells were lysed by sonication in 100 mL buffer (10 mM Tris pH 8.0, 1 mM $CaCl_2$, 1 mM phenylmethysulfonyl fluoride, 1 mM E64, 1 mg/mL pepstatin A, 10 mg/mL soy trypsin inhibitor, 1 mM benzimidine, 1 mg/mL aprotinin, and 1 mg/mL leupeptin). The membrane fraction was collected by centrifugation at 150,000 × *g* for 45 min, and residual nucleic acid was removed from the top of the membrane pellet using Dulbecco's phosphate buffered saline (DPBS). The membrane was homogenized with a Dounce homogenizer in 100 mL lysis buffer (20 mM Tris pH 8.0, 150 mM NaCl, 1 mM $CaCl_2$, 1 mM phenylmethysulfonyl fluoride, 1 mM E64, 1 mg/mL pepstatin A, 10 mg/mL soy trypsin inhibitor, 1 mM benzimidine, 1 mg/mL aprotinin, 1 mg/mL leupeptin, 10 μL Benzonase® endonuclease (EMD Millipore), 1% n-dodecyl-b-D-

maltopyranoside (DDM, Anatrace, Maumee, OH), 0.2% cholesteryl hemisuccinate tris salt (CHS, Anatrace)). Protein was extracted with gentle stirring for 2 h at 4 °C. 5 mL Sepharose resin coupled to anti-GFP nanobody was added to the supernatant and stirred gently for 2 h at 4 °C. The resin was collected in a column and washed with 50 mL buffer 1 (20 mM Tris, 500 mM NaCl, 1 mM $CaCl_2$, 0.025% DDM, 0.005% CHS, pH 8.0), 150 mL buffer 2 (20 mM Tris, 150 mM NaCl, 1 mM $CaCl_2$, 0.025% DDM, 0.005% CHS, pH 8.0), and 20 mL of buffer 1. PPX (~0.5 mg) was added into the washed resin in 5 mL buffer 2 and rocked gently overnight. Cleaved TTYH3 was eluted and concentrated to ~0.5 mL with an Amicon Ultra spin concentrator (50 kDa cutoff, MilliporeSigma, USA). The concentrated protein was subjected to size exclusion chromatography using a Superose 6 Increase 10/300 column (GE Healthcare, Chicago, IL) run in buffer 3 (20 mM Tris pH 8.0, 150 mM NaCl, 1 mM $CaCl_2$, 0.025% DDM, 0.005% CHS) on a NGC system (Bio-Rad, Hercules, CA). The peak fractions were collected and spin concentrated for nanodisc reconstitution[40].

3.5.2 Nanodisc reconstitution

Freshly purified proteins were reconstituted into MSP1E3D1 nanodiscs with a 2:1:1 DOPE:POPS:POPC lipid mixture (mol:mol, Avanti, Alabaster, AL) at a final molar ratio of TTYH:MSP1E3D1:lipid of 1:4:400. Lipids in chloroform were mixed, dried under argon, washed with pentane, dried under argon, and dried under vacuum overnight in the dark. Dried lipids were rehydrated in buffer 4 containing 20 mM Tris pH 8.0, 150 mM NaCl, 1 mM Ca^{2+} and clarified by bath sonication. DDM was added to a final concentration of 8 mM. Proteins were mixed with lipids and incubated at 4 °C for 30 min before addition of MSP1E3D1 protein. After incubation for 10 min at 4 °C, 100 mg of Biobeads SM2 (Bio-Rad, USA) (prepared by sequential washing in methanol, water, and buffer 4 and weighed damp following bulk liquid removal) was added and the mixture was rotated at 4 °C overnight. The sample was spun down to facilitate removal of solution from the Biobeads and reconstituted TTYHs further purified on a Superose 6 increase column. The peak fractions were collected and spin concentrated (50 kDa MWCO) to 1.0–1.3 mg/mL for grid preparation. For cryo-EM samples in absence of Ca^{2+}, 5 mM EGTA was added into the buffer instead of 1 mM $CaCl_2$.

3.5.3 Grid preparation

The TTYH2 and TTYH3 nanodisc samples were centrifuged at 21,000 × *g* for 10 min at 4 °C. A 3 μL sample was applied to holey carbon, 300 mesh R1.2/1.3 gold grids (Quantifoil, Großlöbichau, Germany) that were freshly glow discharged for 25 s. Sample was incubated for 5 s at 4 °C and 100% humidity prior to blotting with Whatman #1 filter paper for 3–3.5 s at blot force 1 and plunge-freezing in liquid ethane cooled by liquid nitrogen using a FEI Mark IV Vitrobot (FEI/Thermo Scientific, USA). For the Ca^{2+}-free condition, 5 mM EGTA was used in place of Ca^{2+} in purification buffers and the filter paper was washed sequentially in water, 5 mM EGTA solution, and water before drying overnight under vacuum to remove the high concentrations of Ca^{2+} in the filter paper that could otherwise be transferred to the sample[29].

3.5.4 Cryo-EM data acquisition

Grids were clipped and transferred to a FEI Talos Arctica electron microscope operated at 200 kV. Fifty frame movies were recorded on a Gatan K3 Summit direct electron detector in super-resolution counting mode with pixel size of 0.5685 Å for TTYH2 and TTYH3 with Ca^{2+} and 0.5775 Å for TTYH2 in absence of Ca^{2+}. The electron dose rate was 9.556, 9.021, and 8.849 e^{-} $Å^2$ s^{-} and the total dose was 50.225, 47.28, and 50.0 e^{-} $Å^2$ for TTYH2 with Ca^{2+}, TTYH3 with Ca^{2+} and TTYH2 without Ca^{2+}, respectively. Nine movies were collected around a central hole position with image shift and defocus was varied from −0.6 to −1.8 μm through SerialEM[41]. See Table 3.1 for data collection statistics.

3.5.5 Cryo-EM data processing

A similar pipeline was followed to process all three datasets[40]. Motion-correction and dose-weighting were performed on all micrograph movies using RELION3.1's implementation of MotionCor2 and 2x "binned" from super resolution to physical pixel size[42,43,44]. CTFFIND-4.1 was used to estimate the contrast transfer function (CTF) parameters[45]. Micrographs were then manually sorted to eliminate subjectively bad micrographs, such as empty or contaminated holes. In addition, micrographs with a CTF maximum estimated resolution lower than 5 Å were discarded. Template-free auto-picking of particles was performed with RELION3.1's Laplacian-of-Gaussian filter yielding an initial set of particles. This initial set of particles underwent classification to generate templates, which were subsequently used for template-based auto-picking of particles. To "clean-up" these particles, they were first 2D-classified in both RELION3.1 and cryoSPARC v2[46,47], then iterative ab initio and heterogenous refinement jobs were used to isolate higher quality particles. After no further improvements in map quality were observed, the resulting subset of particles was used for training in Topaz[48]. The above "clean-up" steps were then repeated for the Topaz-picked particles until no further improvements were observed, at which point the Topaz picks were combined with the template-picked particles, duplicates removed, and input to Bayesian particle polishing in RELION3.1. The resulting "shiny" particles then underwent iterative homogenous and nonuniform refinements until no further improvement was observed. The initial resolution and dynamic mask nonuniform parameters were adjusted to yield the best performing refinement. UCSF pyem was used for conversion of files from cryoSPARC to Relion formats[49].

For the Ca^{2+}-free data, the above strategy was followed without knowledge of the *trans*-dimer state until its discovery during classification of the template-picked particles. After identification and separation of particles into monomeric and *trans*-dimeric classes, Topaz training and subsequent steps of the above pipeline were undertaken independently for each state (Supplemental Figure 3.4).

For the TTYH2 with Ca^{2+} data, C2 symmetry was applied for the final refinement. C2 symmetry was not applied in any other case. We note that a subset of TTYH2 with Ca^{2+} particles yielded an asymmetric TTYH2 *cis*-dimer, similar to the map observed for TTYH3 (Supplemental Figure 3.3). In addition, a small number of monomeric particles were observed, but were insufficient to generate a meaningful reconstruction.

3.5.6 Modeling, refinement, and analysis

Maps were sharpened using cryoSPARC, models were built de novo in Coot[50], real space refined in Phenix[51], and assessed for proper stereochemistry and geometry using Molprobity[52]. Phenix resolve density modification was applied to maps for visualization[53]. The N-terminus prior to TM1 was copied from AlphaFold2 predicted structures[54,55] and refined into the cryo-EM density. Structures were analyzed and figures were prepared with HOLE[56], Dali[27], PyMOL, ChimeraX[57], JalView[58], Prism 8, GNU Image Manipulation Program, and Adobe Photoshop and Illustrator software. Consurf[59] was used to map conservation onto the structure using an alignment of 689 TTYH1-3 sequences retrieved from the Pfam database.

3.5.7 Electrophysiology

We performed whole-cell recordings to study the proposed chloride channel function of TTYH2. The same TTYH2 construct used for structural studies was cotransfected with an AQP4-mCherry construct at a ratio of 1:1 into HEK293T cells with Fugene HD transfection reagent. The cells were grown in DMEM-F12 (Gibco) with 10% FBS, 2 mM L-glutamine, 100 units/mL penicillin and 100 μg/mL streptomycin at 37 °C and 5% CO_2. Whole-cell recordings were performed at room temperature ~24–48 h after transfection. Borosilicate glass pipettes were pulled to a resistance of 3–7 MΩ. An Axopatch 200B amplifier connected to a Digidata 1550B digitizer (Molecular Devices) was used for recording. The following ramp protocol was applied once every 15 s: $V_{hold} = -50$ mV; ramp from +100 mV to −100 mV, 1000 ms. The external bath solutions (isotonic buffer and hypotonic buffer) contained 70 mM Tris-HCl, 1.5 mM $CaCl_2$, 10 mM HEPES, and 10 mM glucose, 5 mM TEA-Cl, and 5 mM BaCl2 adjusted to pH 7.3 with CsOH[22]. The osmolality of each solution was adjusted with sucrose: 100 mM sucrose for ~320 mOsm (for ISO) and 30 mM sucrose for ~240 mOsm (for HOS). Solution osmolality was measured with a vapor pressure osmometer (Vapro #5600, ELITechGroup). During recordings, cells were initially perfused with the isotonic buffer prior to perfusion with hypotonic buffer. The pipette solution contained 60 mM Tris-HCl, 70 mM Tris-Aspartic acid, 15 mM HEPES, 0.4 mM $CaCl_2$, 1 mM $MgCl_2$, 4 mM Mg-ATP, 0.5 mM Na-GTP, and 1 mM EGTA adjusted to pH 7.25 by CsOH. All data were acquired using Clampex 10.7 (Molecular Devices) and analyzed in Clampfit 10.7 (Molecular Devices) and GraphPad prism.

3.6 Data availability

All data associated with this study will be publicly available. For the TTYH2 *cis*-dimer in the presence of Ca^{2+}, the final model is in the PDB under 7RTT, the final map is in the Electron Microscopy Data Bank (EMDB) under EMD-24688, and the original micrograph movies and final particle stack are in the Electron Microscopy Public Image Archive (EMPIAR) database under EMPIAR-10842. For the TTYH2 *trans*-dimer in the absence of Ca^{2+}, the final model is in the PDB under 7RTU and the final map is in the EMDB under EMD-24689. For the TTYH2 monomer in the absence of Ca^{2+}, the final model is in the PDB under 7RTV and the final map is in the EMDB under EMD-24690. Original micrograph movies and final particle stacks for the TTYH2 monomer and *trans*-dimer are in the EMPIAR database under EMPIAR-10843. For the TTYH3 *cis*-dimer in the presence of Ca^{2+}, the final model is in the PDB under 7RTW, the final map is in the EMDB under EMD-24691, and the original micrograph movies and final particle stack are in the EMPIAR database under EMPIAR-10850. Source data are available online (https://www.nature.com/articles/s41467-021-27283-8).

3.7 Acknowledgements

The authors thank J. Remis, D. Toso, and P. Tobias for microscope and computational support at the Cal-Cryo facility. The authors thank members of the Brohawn Laboratory for discussions and feedback on the project. S.G.B. is a New York Stem Cell Foundation-Robertson Neuroscience Investigator. This work was funded by the New York Stem Cell Foundation; NIGMS grant no. GM123496; a McKnight Foundation Scholar Award; a Sloan Research Fellowship; and a Winkler Family Scholar Award (to S.G.B.).

3.8 Contributions

B.L. generated expression constructs, purified proteins, prepared samples for cryo-EM, and performed electrophysiology experiments. B.L. and C.M.H. collected cryo-EM data, processed cryo-EM data, and determined and modeled structures. All authors analyzed data and wrote the manuscript. S.G.B. supervised the project.

3.9 References

1. Campbell, H. D. et al. Human and mouse homologues of the *Drosophila melanogaster* tweety (tty) gene: a novel gene family encoding predicted transmembrane proteins. *Genomics* **68**, 89–92 (2000).
2. Rae, F. K. et al. TTYH2, a human homologue of the *Drosophila melanogaster* gene tweety, is located on 17q24 and upregulated in renal cell carcinoma. *Genomics* **77**, 200–207 (2001).
3. Suzuki, M. & Mizuno, A. A novel human Cl(-) channel family related to *Drosophila* flightless locus. *J. Biol. Chem.* **279**, 22461–22468 (2004).
4. Matthews, C. A. et al. Expression and evolution of the mammalian brain gene Ttyh1. *J. Neurochem.* **100**, 693–707 (2007).
5. Nalamalapu, R. R., Yue, M., Stone, A. R., Murphy, S. & Saha, M. S. The tweety gene family: from embryo to disease. *Front. Mol. Neurosci.* **14**, 672511 (2021).
6. Campbell, H. D. et al. The *Drosophila melanogaster* flightless-I gene involved in gastrulation and muscle degeneration encodes gelsolin-like and leucine-rich repeat domains and is conserved in *Caenorhabditis elegans* and humans. *Proc. Natl Acad. Sci. USA* **90**, 11386–11390 (1993).
7. Maleszka, R., Hanes, S. D., Hackett, R. L., Couet, H. Gde & Miklos, G. L. The *Drosophila melanogaster* dodo (dod) gene, conserved in humans, is functionally interchangeable with the ESS1 cell division gene of Saccharomyces cerevisiae. *Proc. Natl Acad. Sci. USA* **93**, 447–451 (1996).
8. Kumada, T. et al. Ttyh1, a Ca2+-binding protein localized to the endoplasmic reticulum, is required for early embryonic development. *Dev. Dynam.* **239**, 2233–2245 (2010).
9. Halleran, A. D. et al. Characterization of tweety gene (ttyh1-3) expression in *Xenopus laevis* during embryonic development. *Gene Expr. Patterns* **17**, 38–44 (2015).
10. Kleinman, C. L. et al. Fusion of TTYH1 with the C19MC microRNA cluster drives expression of a brain-specific DNMT3B isoform in the embryonal brain tumor ETMR. *Nat. Genet.* **46**, 39–44 (2014).

11. Toiyama, Y. TTYH2, a human homologue of the *Drosophila melanogaster* gene tweety, is up-regulated in colon carcinoma and involved in cell proliferation and cell aggregation. *World J. Gastroenterol.* **13**, 2717 (2007).
12. Moon, D. K., Bae, Y. J., Jeong, G.-R., Cho, C.-H. & Hwang, S. C. Upregulated TTYH2 expression is critical for the invasion and migration of U2OS human osteosarcoma cell lines. *Biochem. Biophys. Res. Commun.* **516**, 521–525 (2019).
13. Jung, E. et al. Tweety-homolog 1 drives brain colonization of gliomas. *J. Neurosci.* **37**, 6837–6850 (2017).
14. Saha, S. K., Biswas, P. K., Gil, M. & Cho, S.-G. High expression of TTYH3 is related to poor clinical outcomes in human gastric cancer. *J. Clin. Med.* **8**, 1762 (2019).
15. Weinberg, F. et al. Identification and characterization of a BRAF fusion oncoprotein with retained autoinhibitory domains. *Oncogene* **39**, 814–832 (2020).
16. Lu, Q. et al. SARS-CoV-2 exacerbates proinflammatory responses in myeloid cells through C-type lectin receptors and tweety family member 2. *Immunity* **54**, 1304–1319.e9 (2021).
17. Stefaniuk, M., Swiech, L., Dzwonek, J. & Lukasiuk, K. Expression of Ttyh1, a member of the tweety family in neurons in vitro and in vivo and its potential role in brain pathology. *J. Neurochem.* **115**, 1183–1194 (2010).
18. Wiernasz, E. et al. Ttyh1 protein is expressed in glia in vitro and shows elevated expression in activated astrocytes following status epilepticus. *Neurochem. Res.* **39**, 2516–2526 (2014).
19. Han, W.-J. et al. Tweety-homolog 1 facilitates pain via enhancement of nociceptor excitability and spinal synaptic transmission. *Neurosci. Bull.* 1–19 (2020). https://doi.org/10.1007/s12264-020-00617-0.
20. Reyes, J. et al. Delimitation of the pore in tweety homolog 1 channels: a model-guided approach. *J. Adv. Biol. Biotechnol.* **10**, 1–12 (2016).
21. Ryu, J. et al. Surface expression of TTYH2 is attenuated by direct interaction with β-COP. *BMB Rep.* **52**, 445–450 (2019).
22. Han, Y.-E. et al. Tweety-homolog (Ttyh) family encodes the pore-forming subunits of the swelling-dependent volume-regulated anion channel (VRACswell) in the brain. *Exp. Neurobiol.* **28**, 183–215 (2019).
23. Bae, Y. et al. TTYH1 and TTYH2 serve as LRRC8A-independent volume-regulated anion channels in cancer cells. *Cells* **8**, 562–14 (2019).
24. Suzuki, M. The *Drosophila* tweety family: molecular candidates for large-conductance Ca2+-activated Cl– channels. *Exp. Physiol*. 141–147 (2006). https://doi.org/10.1113/expphysiol.2005.031773.
25. Kim, J. et al. Ttyh1 regulates embryonic neural stem cell properties by enhancing the Notch signaling pathway. *EMBO Rep.* **19**, e45472 (2018).
26. Wu, H.-N. et al. Deficiency of Ttyh1 downstream to Notch signaling results in precocious differentiation of neural stem cells. *Biochem. Biophys. Res. Commun.* **514**, 842–847 (2019).
27. Holm, L. & Sander, C. Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480 (1995).
28. He, Y. et al. N-glycosylation analysis of the human tweety family of putative chloride ion channels supports a penta-spanning membrane arrangement: impact of N-glycosylation on cellular processing of tweety homologue 2 (TTYH2). *Biochem. J.* **412**, 45–55 (2008).

29. Paknejad, N. & Hite, R. K. Structural basis for the regulation of inositol trisphosphate receptors by Ca2+ and IP3. *Nat. Struct. Mol. Biol.* **25**, 660–668 (2018).
30. Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T. & MacKinnon, R. X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature* **415**, 287–294 (2002).
31. Paulino, C., Kalienkova, V., Lam, A. K. M., Neldner, Y. & Dutzler, R. Activation mechanism of the calcium-activated chloride channel TMEM16A revealed by cryo-EM. *Nature* **552**, 421–425 (2017).
32. Dang, S. et al. Cryo-EM structures of the TMEM16A calcium-activated chloride channel. *Nature* **552**, 426–429 (2017).
33. Okada, Y. et al. The Puzzles of Volume-Activated Anion Channels. *Physiology and pathology of chloride transporters and channels in the nervous system Ch. 15.* 283–306 (Academic Press, 2010). https://doi.org/10.1016/b978-0-12-374373-2.00015-7.
34. Sabirov, R. Z. et al. The ATP-releasing maxi-Cl channel: its identity, molecular partners, and physiological/pathophysiological implications. *Life* **11**, 509 (2021).
35. He, Y. et al. The ubiquitin-protein ligase Nedd4-2 differentially interacts with and regulates members of the tweety family of chloride ion channels. *J. Biol. Chem.* **283**, 24000–24010 (2008).
36. Nagae, M. et al. Crystal structure of α5β1 integrin ectodomain: atomic details of the fibronectin receptor. *J. Cell Biol.* **197**, 131–140 (2012).
37. Ludwig, B. S., Kessler, H., Kossatz, S. & Reuning, U. RGD-binding integrins revisited: how recently discovered functions and novel synthetic ligands (re-)shape an ever-evolving field. *Cancers* **13**, 1711 (2021).
38. Sukalskaia, A., Straub, M. S., Deneka, D., Sawicka, M. & Dutzler, R. Cryo-EM structures of the TTYH family reveal a novel architecture for lipid interactions. *Nat. Commun.* **12**, 4893 (2021).
39. Goehring, A. et al. Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. *Nat. Protoc.* **9**, 2574–2585 (2014).
40. Kern, D. M. et al. Cryo-EM structure of SARS-CoV-2 ORF3a in lipid nanodiscs. *Nat. Struct. Mol. Biol.* 1–10 (2021). https://doi.org/10.1038/s41594-021-00619-0.
41. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
42. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
43. Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5–17 (2019).
44. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, 163 (2018).
45. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
46. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
47. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 (2020).
48. Bepler, T. et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019).

49. Asarnow, D., Palovcak, E., Cheng, Y. UCSF pyem v0.5. *Zenodo* https://doi.org/10.5281/zenodo.3576630 (2019).
50. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D, Biol. Crystallogr.* **66**, 486–501 (2010).
51. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. Sect. D, Struct. Biol.* **75**, 861–877 (2019).
52. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
53. Terwilliger, T. C., Ludtke, S. J., Read, R. J., Adams, P. D. & Afonine, P. V. Improvement of cryo-EM maps by density modification. *Nat. Methods* **17**, 923–927 (2020).
54. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 1–11 (2021). https://doi.org/10.1038/s41586-021-03819-2.
55. Tunyasuvunakool, K. et al. Highly accurate protein structure prediction for the human proteome. *Nature* 1–9 (2021). https://doi.org/10.1038/s41586-021-03828-1.
56. Smart, O. S., Neduvelil, J. G., Wang, X., Wallace, B. A. & Sansom, M. S. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* **14**, 354-60–354376 (1996).
57. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
58. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
59. Ashkenazy, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).

3.10 Tables

	TTYH2 cis-dimer with Ca^{2+} (PDB 7RTT) (EMDB 24688)	TTYH3 cis-dimer with Ca^{2+} (PDB 7RTW) (EMDB 24691)	TTYH2 trans-dimer without Ca^{2+} (PDB 7RTU) (EMDB 24689)	TTYH2 monomer without Ca^{2+} (PDB 7RTV) (EMDB 24690)
Data collection and processing				
Magnification	36,000 x	36,000 x	36,000 x	165,000 x
Voltage (kV)	200	200	200	200
Micrographs (no.)	5014	3217	7245	7245
Electron exposure (e^{-}/Å^2)	50.225	47.28	50	50
Defocus range (μm)	-0.6 to -1.8	-0.6 to -1.8	-0.6 to -1.8	-0.6 to -1.8
Super resolution pixel size (Å)	0.5685	0.5685	0.5575	0.5575
Map pixel size (Å)	1.137	1.137	1.115	1.699
Symmetry imposed	C2	C1	C1	C1
Topaz particle images (no.)	4,220,298	1,475,741	1,380,683	2,351,846
Final particle images (no.)	96,342	96,285	135,683	186,697
Map resolution (Å)	3.34	3.23	3.89	3.96
FSC threshold	0.143	0.143	0.143	0.143
Refinement				
Initial model used (PDB code)	*de novo*	*de novo*	7RTT	7RTT
Model resolution (Å)	3.5	3.5	3.8	4.1
FSC threshold	0.143	0.143	0.143	0.143
Model composition				

Nonhydrogen atoms	6,226	5,630	5,960	2,880
Protein residues	791	699	754	366
Ligands	14	17	12	6
B factors (Å^2)				
Protein	80.27	57.43	107.75	108.17
Ligand	96.52	57.70	110.82	193.51
R.m.s. deviations				
Bond lengths (Å)	0.003	0.003	0.005	0.003
Bond angles (°)	0.544	0.566	0.692	0.727
Validation				
MolProbity score	1.6	1.74	1.78	1.77
Clashscore	5.26	9.68	9.63	13.18
Poor rotamers (%)	0	0	1.26	0
Ramachandran plot				
Favored (%)	95.40	96.53	96.78	97.24
Allowed (%)	4.6	3.47	3.22	2.76
Disallowed (%)	0	0	0	0

Table 3.1: Cryo-EM data collection, refinement, and validation statistics.

mTTYH2:

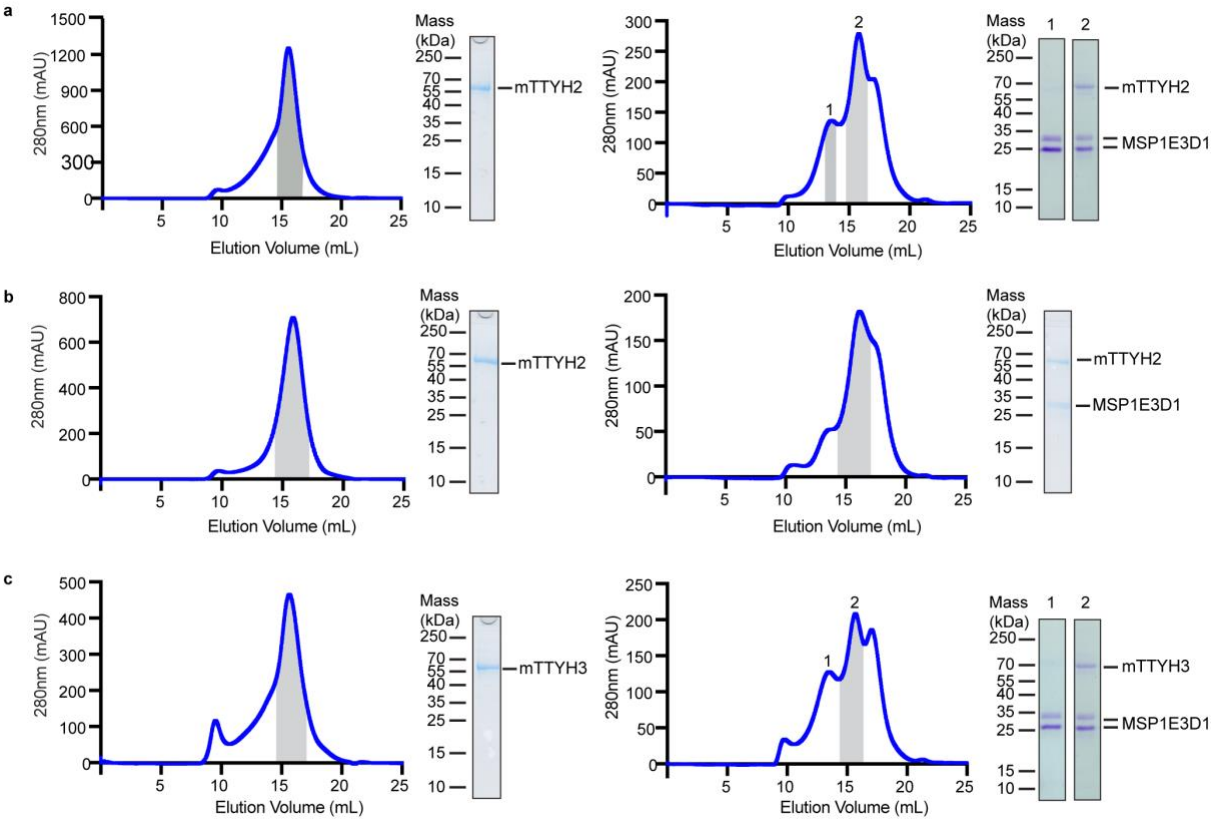
ATGCCAGCTGCACGCGTGGAATATATCGCACCTTGGTGGGTTGTGTGGTTGCACTCCGTTCCTCATTT
GGGCCTGCGACTGCAGAGAGTGGATTCTACCTTTAGTCCTGGAGACGAAACCTATCAAGAGAGCCTC
CTCTTCTTGGGAGTACTCGCAGCCATTGGTCTCGGATTGAATCTTATATTTTTGACCGTATATCTCGTTT
GCACTTGCTGTTGTAGGCGCGATCACACAGTTCAAACAAAACAACAGGAAAGCTGCTGTGTTACCTG
GACTGCTGTGGTGGCTGGCCTGCTGTGTTGCGCAGCTGTCGGGGTAGGATTCTATGGAAACTCAGAA
ACCAACGACGGGATGCACCAACTCATATATAGTCTGGACAACGCTAACCACACATTCAGCGGTATGG
ATGAATTGGTTTCAGCAAACACTCAGAGGATGAAAGTCGATCTTGAACAACATCTCGCTAGGCTGTCA
GAGATCATAGCAGCAAGGGGCGATTATATCCAAACTCTCAAGTTTATGCAGCAGATGGCAGGGAACG
TTGTTTCACAGTTGTCCGGCTTGCCCGTGTGGAGAGAGGTCACCACCCAGTTGACCAAGCTGTCTCAT
CAAACAGCATACGTTGAGTACTATAGGTGGCTTAGTTACCTTCTTCTTTTTATACTTGATCTCGTCATTT
GTCTCGTCACTTGCCTGGGGCTGGCCCGACGCTCAAAATGTCTGTTGGCCAGCATGCTGTGTTGCGGG
ATACTGACATTGATTTTGAGCTGGGCTTCTCTGGCCGCCGACGCTGCCGCCGCTGTGGGGACTTCTGA
CTTTTGCATGGCCCCCGATATATACATTCTCAATAACACTGGGTCTCAGATTAACTCCGAGGTCACTCG
GTATTACCTTCACTGCTCTCAAAGTCTTATTAGCCCCTTTCAACAATCTCTGACTACATTCCAAAGGAGT
TTGACAACTATGCAGATTCAGGTGGGGGGCTTGCTTCAATTTGCTGTCCCCCTTTTTCCAACCGCAGA
GAAGGACCTGTTGGGCATTCAACTCTTGTTGAATAACTCTGAAATCTCTCTGCACCAGCTCACCGCAAT
GCTTGATTGCCGGGGTTTGCACAAGGATTATCTCGATGCACTGACTGGAATTTGTTATGATGGCATCG
AGGGCTTGCTCTTCCTCGGGCTTTTCTCACTGCTCGCTGCATTGGCCTTTAGTACACTGACTTGCGCAG
GCCCTAGAGCCTGGAAATATTTCATAAATCGCGACAGGGATTATGACGATATAGATGATGATGACCC
CTTTAACCCACAGGCCAGGCGCATCGCCGCACACAATCCTACCAGAGGGCAACTTCATTCCTTTTGTTC
TTACTCCTCCGGACTCGGATCACAATGCTCTTTGCAGCCTCCTTCTCAGACCATCAGTAATGCTCCTGTC
TCCGAGTATATGAATCAAGCCATCCTTTTCGGGGGAAATCCACGATACGAAAACGTACCACTTATTGG
TCGAGGATCCCCTCCCCCTACTTACAGCCCTAGCATGAGACCCACCTATATGTCTGTCGCAGACGAAC
ATTTGCGCCACTACGAGTTCCCCTCC

mTTYH3:

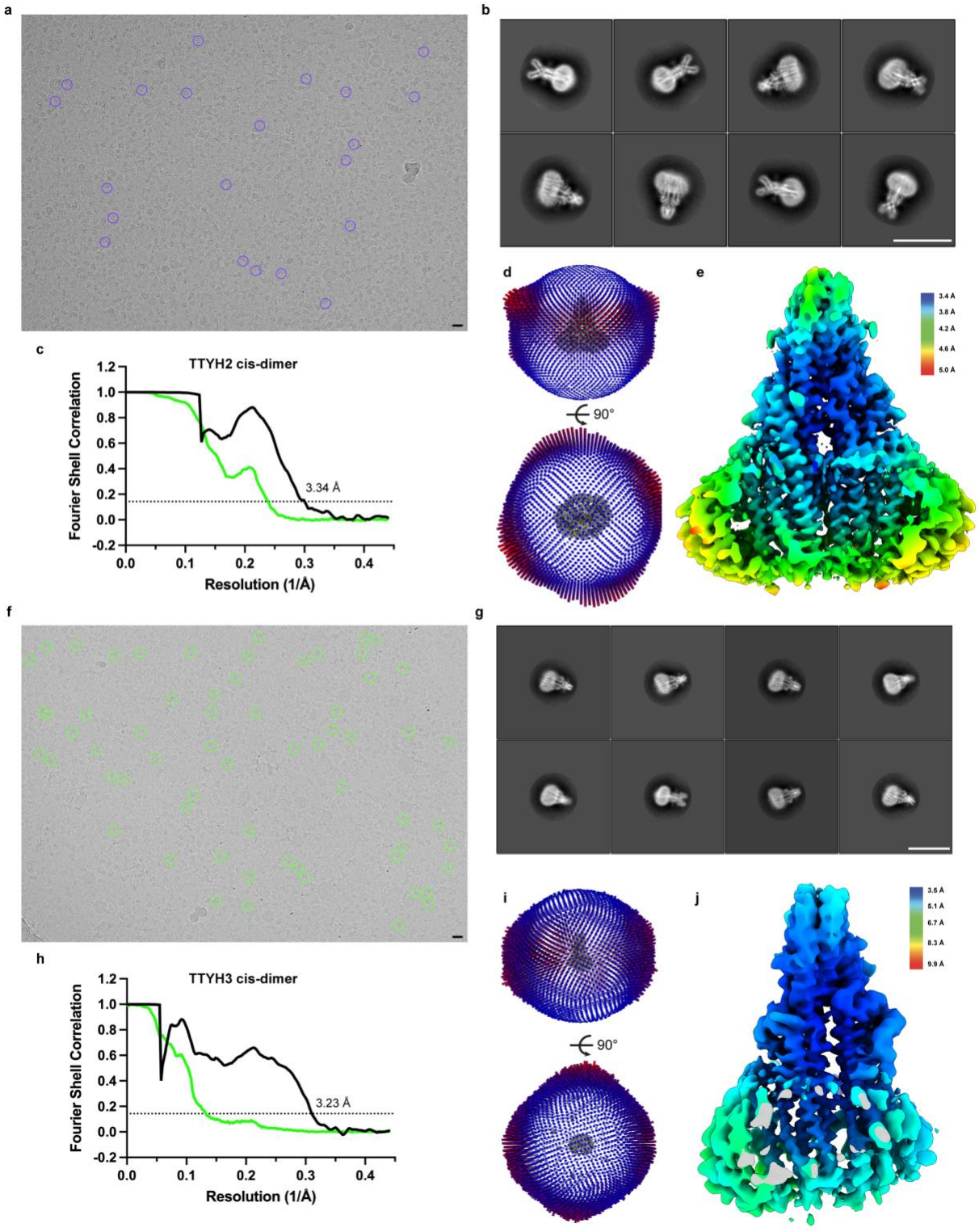
ATGGCAGGCGTAAGCTATGCTGCCCCTTGGTGGGTCTCACTTCTCCATCGCCTTCCCCACTTTGACCTG
CGGTGGGAAGCCACTAGCTCACAATTTAGACCTGAGGACGCAGACTATCAGCAAGCACTCCTTCTTTT
GGGGGCCACTGCCCTTGCTTGTCTGGCACTGGATTTGCTGTTCCTCTTGTTCTACTCTTTCTGGCTTTGC
TGCCGAAGGAGAAAAACAGACGAACACTTGGACGCCGACTGTTGTTGTACCGCCTGGTGCGTGATTA
TCACCACATTGGTGTGTTCAGCCGGGATTGCAGTTGGTTTCTATGGGAATGGGGAAACCTCTGATGGT
ATCCACCGAGCTACATACAGCCTGCGGCACGCTAATAGAACAGTAGCTGGGGTACAAGATCGAGTCT
GGGATACAGCCGCTGCCCTCAATCGAACCGCCGAGCCCAACCTGCAGTCTCTTGAAAGACAGCTCGC
CGGACGCCAGGAACCCCTGCGGGCTGTGCAACGCCTTCAAACTCTCCTTGGTACTCTTCTCGGGTATA
CAGCCGCTATCCCCTTCTGGAGGAATCCAGGTGTTAGCTTGGAAGTCCTTGCTGAGCAAGTGGATTTG
TATGACTGGTACAGATGGCTGGGGTACTTGGGTCTGCTTCTCCTGGATGTGATTATTTGCCTTCTTGTG
CTTGTAGGATTGATAAGGTCATCAAAAGGCATTCTTGTAGGCGTCTGTTTGTTGGGCGTACTGGCCTT
GGTTATATCATGGGGAGCCTTGGGATTGGAACTGGCTGTCAGTGTGGGCTCTAGCGATTTCTGCGTT
GACCCTGATACTTTTGTGACAAAAATGGTGGAAGAGCATAGTGTGCTCTCTGGGGATATTTTGCAGTA
TTATCTGGCCTGCTCCCCCCGAGCCACTAATCCATTCCAACAGAAATTGTCTGGTTCACATAAAGCACT
GGTAGAAATGCAGGATGTGGTCGCTGAGCTGCTCAGGAATGTGCCCAGGGAGCACCCAGCTACAAA
AGACCCCCTCCTTAGAGTCCAAGAAGTCCTGAATGGTACAGAAGTCAATCTCCAGCATCTGACTGCCC
TCGTAGATTGCCGATCTTTGCATCTTGACTACGTACAGGCTTTGACAGGCTTTTGTTACGATGGCGTTG
AAGGTTTGATCTATCTTGCACTGTTCAGCTTTGTGACAGCCTTGATGTTCTCCAGTATCGTCTGCAGTA
TTCCACATACTTGGCAACAAAAACGGGGCCCTGACGATGACGGAGAAGAAGAAACCGCACCAGGAC
CTAGGCAAGCTCACGACTCACTTTATAGAGTCCACATGCCTTCACTGTACAGCTGTGGATCTAGCTAC
GGTAGCGAGGCATCCATACCTGCCGCTGCACACACAGTTAGCAATGCTCCCGTTACAGAGTACATGTC
TCAGAACGCCAATTTTCAGAACCCTCGATGCGAGAATACACCTCTCATTGGCCGCGAATCACCTCCCCC
ATCCTATACAAGTTCCATGCGAGCAAAATACTTGGCAACTTCCCAACCCAGGCCCGATTCTAGCGGAT
CTGGCCAC

Table 3.2: Codon optimized gene sequences

3.11 Supplemental figures

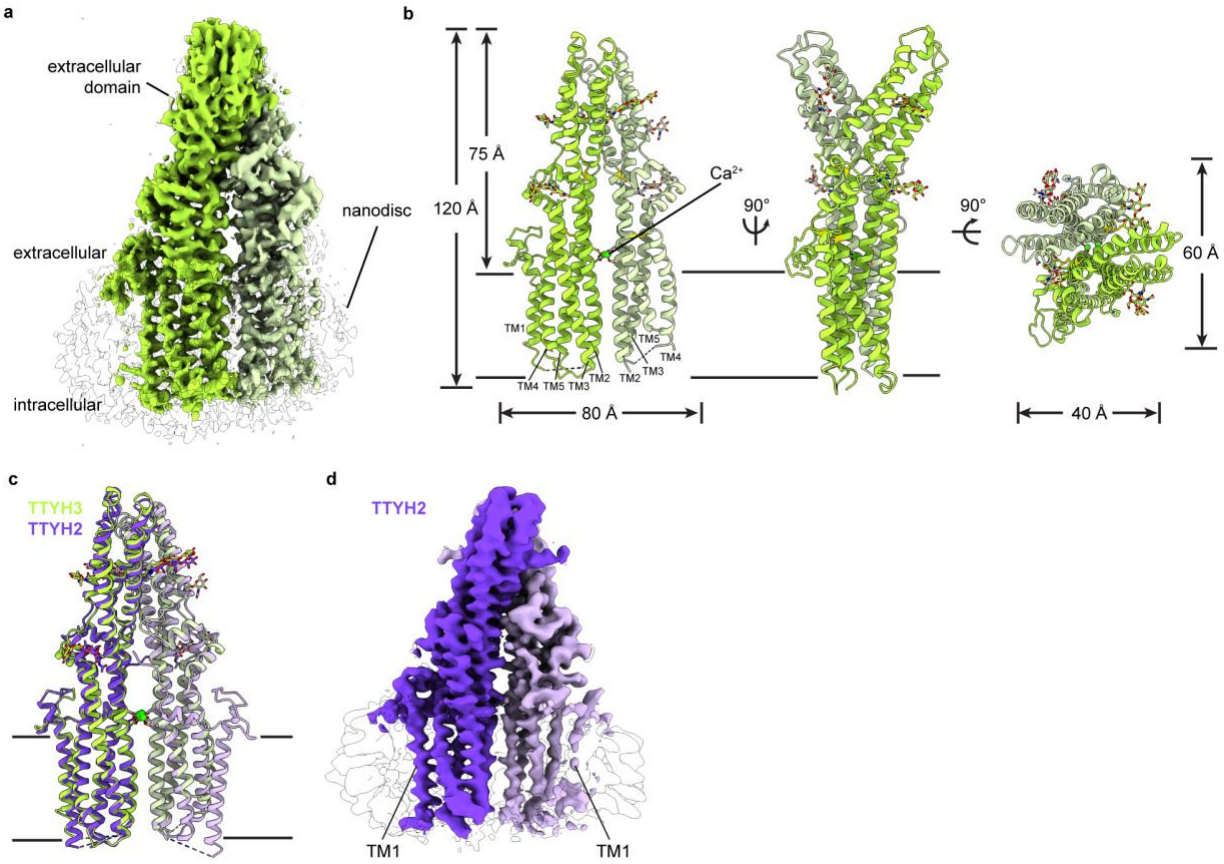


Supplemental Figure 3.1: Purification and lipid nanodisc reconstitution of TTYHs.
a, Purification and reconstitution of TTYH2 in the presence of 1 mM Ca 2+ . Size exclusion chromatogram of TTYH2 purified in DDM/CHS detergent and 1 mM Ca2+ (left) and reconstituted in MSP1E3D1 lipid nanodiscs (right). Pooled fractions are highlighted with corresponding Coomassie-stained SDS-PAGE of the final sample is shown. Early eluting peak 1 corresponds to TTYH2-free MSP1E3D1 containing liposomes or aggregates. **b**, Purification and reconstitution of TTYH2 in the absence of Ca 2+ . Size exclusion chromatogram of TTYH2 purified in DDM/CHS detergent (left) and reconstituted in MSP1E3D1 lipid nanodiscs (right). Pooled fractions are highlighted with corresponding Coomassie-stained SDS-PAGE of the final sample is shown. **c,** Purification and reconstitution of TTYH3 in the presence of 1 mM Ca 2+ . Size exclusion chromatogram of TTYH3 purified in DDM/CHS detergent and 1 mM Ca2+ (left) and reconstituted in MSP1E3D1 lipid nanodiscs (right). Pooled fractions are highlighted with corresponding Coomassie-stained SDS-PAGE of the final sample is shown. Early eluting peak 1 corresponds to TTYH3-free MSP1E3D1 containing liposomes or aggregates. Samples were run on a Superose 6 column.

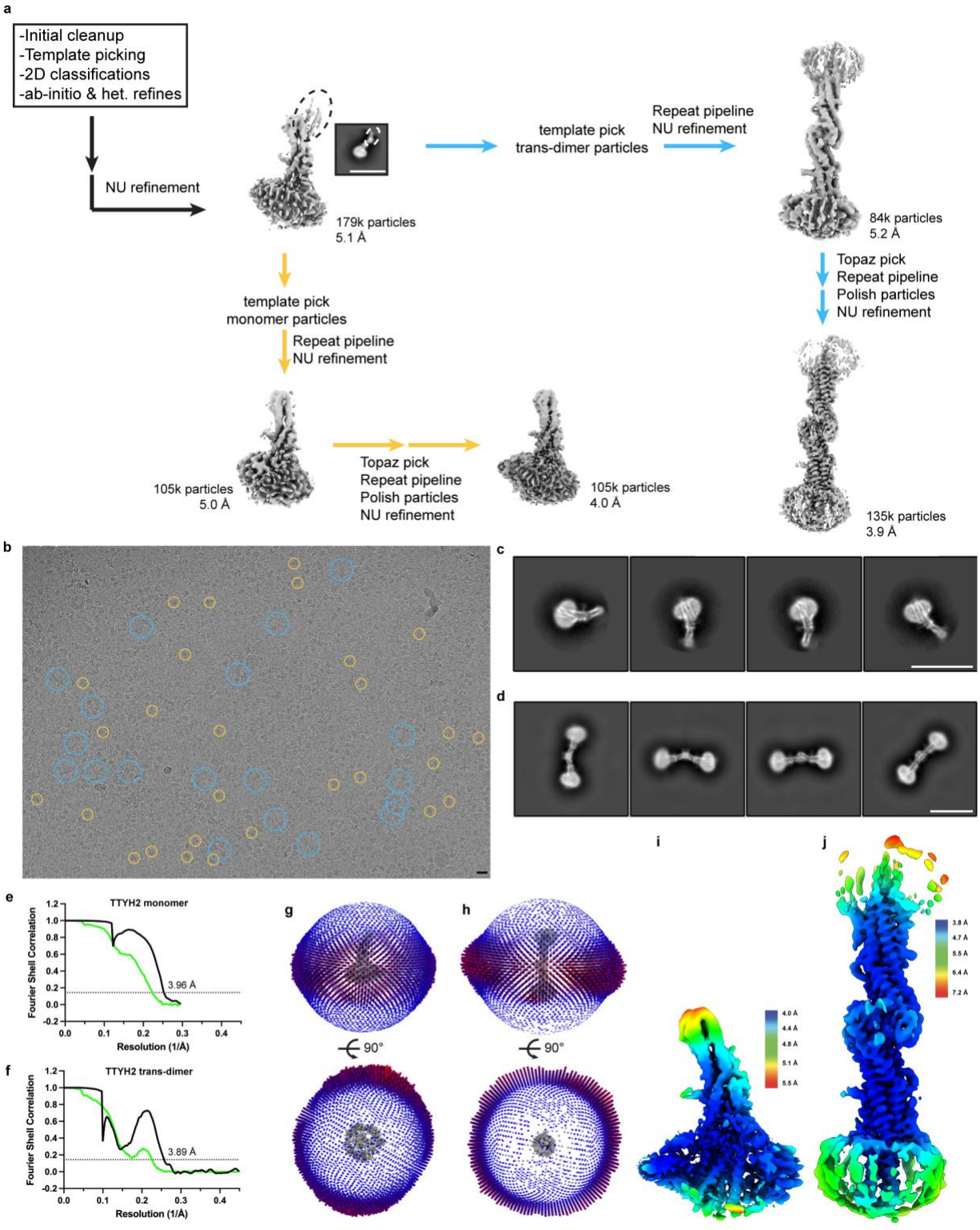


Supplemental Figure 3.2: Example micrographs, 2D class averages, and validation for TTYH2 and TTYH3 cis-dimers in the presence of Ca2+. a,f, Representative micrographs (of 5,014 and 3,217 total micrographs, respectively, **b,g,** selected 2D class averages, **c,h,** Fourier shell correlations between the two corrected (black) and unmasked (green) half maps, **d,i,** angular distribution of particles used in final refinement and **e,j,** Relion-estimated

local resolution colored on the final maps of **a-e** TTYH2 and **f-j** TTYH3 cis-dimers determined in the presence of 1 mM Ca 2+. Scale bars are 150 Å.

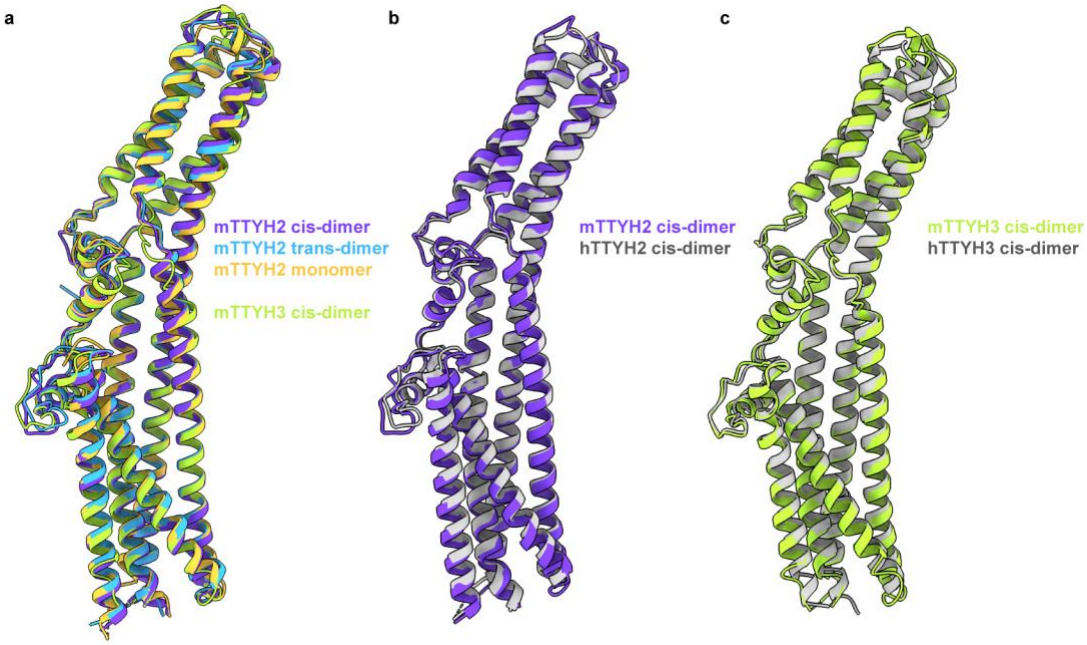


Supplemental Figure 3.3: Structure of a TTYH3 cis-dimer in the presence of Ca 2+ . a, Cryo-EM map of a TTYH3 cis-dimer in MSP1E3D1 nanodiscs at 3.2 Å resolution viewed from the membrane plane. Density from one TTYH3 protomer is colored dark green, the second protomer is light green, and the nanodisc is white. **b,** Model of the TTYH3 cis-dimer viewed from the membrane and from the extracellular side. N-linked glycosylation sites and disulfide bonds in the extracellular domain are drawn as sticks and Ca 2+ ions are shown as green spheres. **c,** Overlay of TTYH3 (greens) and TTYH2 (purples) cis-dimer models. TM1 is modeled in the both TTYH2 protomers, but is not modeled in the light green TTYH3 protomer. **d,** Cryo-EM reconstruction from a subset of TTYH2 particles with an asymmetric nanodisc in which density corresponding to TM1 from the light purple protomer is poorly resolved and the neighboring disc density is weaker as observed in the TTYH3 reconstruction.

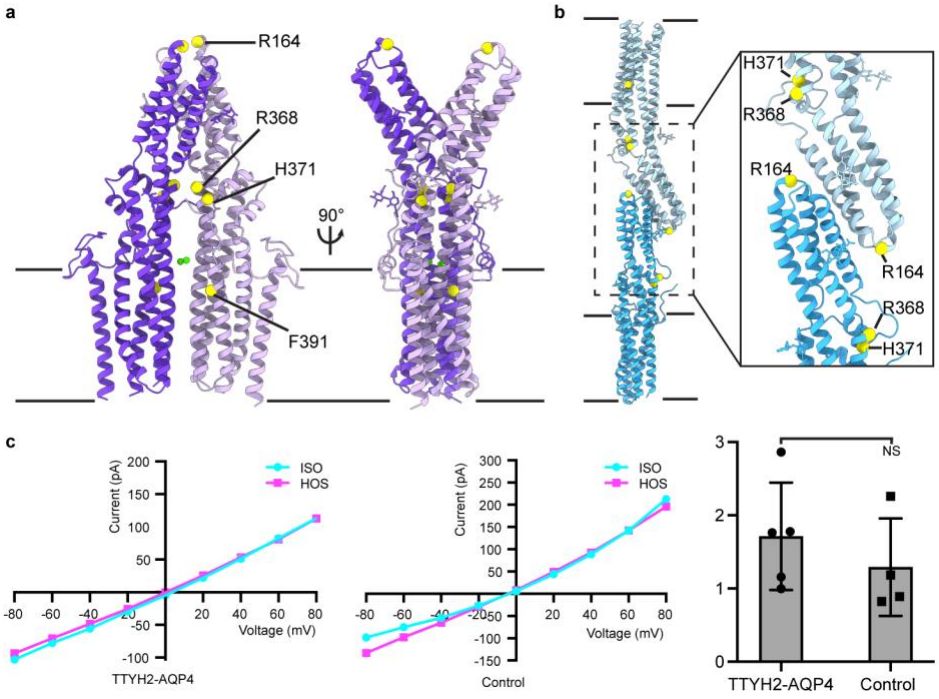


Supplemental Figure 3.4: Cryo-EM data processing and validation for TTYH2 monomers and trans-dimers in the absence of Ca2+. a, Overview of data processing pipeline in cryoSPARC and Relion. Extra density at the distal end of the extracellular domain observed early in processing (circled with dashed lines) suggested the presence of trans-dimers in the dataset. **b,** Representative micrograph (of 7245 total micrographs) with trans-dimers circled in blue and monomers circled in orange. **c,** Selected 2D class averages of monomers and **d,** trans-dimers. **e,f,**

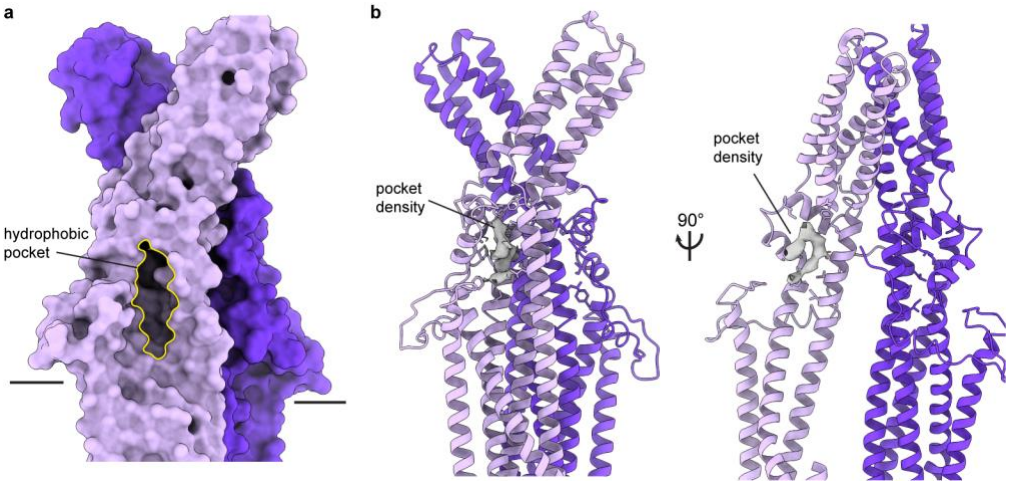
Fourier shell correlations between the two corrected (black) and unmasked (green) half maps. **g,** angular distribution of particles used in final monomer and **h,** trans-dimer refinements. **i,** Relion-estimated local resolution colored on the final monomer and **j,** trans-dimer maps. Scale bars are 150 Å.



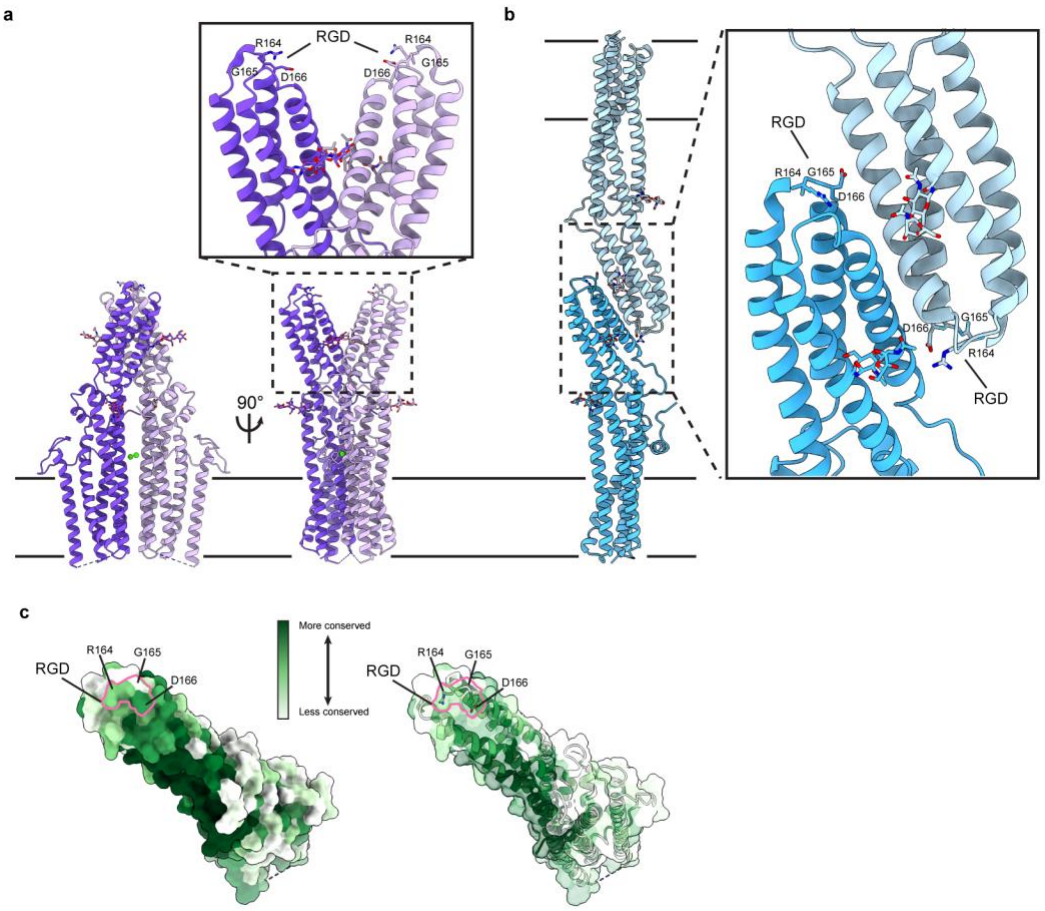
Supplemental Figure 3.5: Comparison of TTYH protomers from different structures. a, Overlay of the mTTYH2 monomer (orange, PDB: 7RTV; r.m.s.d. = 1.3 Å), one protomer from the mTTYH2 cis-dimer (purple, PDB: 7RTT), one protomer from the mTTYH3 cis-dimer (green, PDB: 7RTW; r.m.s.d. = 1.4 Å), and one protomer from the mTTYH2 trans-dimer (blue, PDB: 7RTU; r.m.s.d. = 1.5 Å). r.m.s.d.s are between protomers of the indicated structure and the mTTYH2 cis-dimer. **b,** Overlay of one protomer from the mTTYH2 cis-dimer (purple, PDB: 7RTT) and hTTYH2 cis-dimer (gray, PDB: 7P54; r.m.s.d. = 1.2 Å). **c,** Overlay of one protomer from the mTTYH3 cis-dimer (purple, PDB: 7RTW) and hTTYH3 cis-dimer (gray, PDB: 7P5C; r.m.s.d. = 1.2 Å). Large scale conformational changes are not observed between the structures.



Supplemental Figure 3.6: Mutations previously reported to impact channel activity and lack of TTYH2-dependent currents in heterologous cells. a, TTYH2 cis-dimer and **b,** TTYH2 trans-dimer viewed from the membrane plane with positions of point mutations previously reported to impact TTYH-dependent channel activity indicated with yellow spheres. **c,** Representative current-voltage relationships and maximum fold-activation at 80 mV recorded in response to hypotonic extracellular solution-induced cell swelling (HOS). No significant difference between TTYH2 and AQP4 co-expressing cells and control cells is observed (mean ± sem, n=5 and 4 cells for TTYH2/AQP4 co-expressing and control cells, respectively, two-tailed p=0.41, unpaired Student's t-test).



Supplemental Figure 3.7: A membrane-proximal hydrophobic pocket in the extracellular domain of TTYHs. a, TTYH2 cis-dimer molecular surface shown from the membrane plane. The hydrophobic pocket is outlined in yellow. **b,** TTYH2 model with unassigned horseshoe-shaped pocket density indicated and hydrophobic residues lining the pocket shown as sticks



Supplemental Figure 3.8: An RGD motif at the distal tip of the TTYH2 extracellular domain. a, TTYH2 cis-dimer and **b,** trans-dimer with the RGD motifs within the ED1-ED2 linker drawn as sticks. The RGD motifs are solvent-exposed in cis-dimers and monomers and partially occluded in trans-dimers at the interface between protomers. **c**, TTYH2 surface viewed from above and colored according to conservation among chordate TTYH1-3 sequences. An opaque surface (left) and transparent surface view (right) are shown with the RGD motif outlined in pink and position of each residue indicated.

Chapter 4 Cryo-EM structure of the GOLD-domain seven-transmembrane protein TMEM87A

4.1 Abstract

TMEM87s are eukaryotic transmembrane proteins with two members (TMEM87A and TMEM87B) in humans. TMEM87A and TMEM87B have proposed roles in protein transport to and from the Golgi, as mechanosensitive ion channels, and in developmental signaling. TMEM87 disruption has been implicated in cancers and developmental disorders. To better understand TMEM87 structure and function, we determined a cryo-EM structure of human TMEM87A in lipid nanodiscs. TMEM87A consists of a Golgi dynamics (GOLD) domain atop a membrane spanning seven-transmembrane helix domain containing a large cavity exposed to the extracellular or lumenal space. Structural and functional analyses suggest TMEM87A may not function as an ion channel or G-protein coupled receptor. We find TMEM87A shares its characteristic domain arrangement with seven additional proteins in humans; four previously identified as evolutionary related (TMEM87A, TMEM87B, GPR107, and GPR108) and four additional structural homologs (GPR180, TMEM145, TMEM181, and WLS). Within this group, WLS has been best characterized as a membrane trafficking and secretion chaperone for lipidated Wnt signaling proteins. We suggest TMEM87A could similarly bind lipidated cargo and GOLD domain seven-transmembrane helix (GOST family) proteins could constitute an evolutionarily related family with roles in membrane trafficking of hydrophobic cargo.

4.2 Introduction

TMEM87 proteins are transmembrane proteins found in eukaryotic organisms from fungi and plants to mammals.[1] In humans two paralogs have been identified: TMEM87A and TMEM87B.[1] TMEM87A and TMEM87B have been implicated in regulation of protein transport to and from the Golgi, mechanosensitive cation channel activity, and developmental processes including cardiac development.[2-5] In humans, TMEM87 proteins have been implicated in diseases such as cancer and developmental disease.[6-9]

Several lines of evidence support a role for TMEM87 in aspects of intracellular protein trafficking. First, TMEM87s localize to the Golgi.[2-3] Second, overexpression of TMEM87A or TMEM87B partially rescues defective endosome-to-*trans*-Golgi network retrograde traffic observed in HEK293 cells lacking the Golgi-associated retrograde protein (GARP) complex member VPS54.[2] Third, TMEM87A has been identified as retrograde cargo captured by mitochondrial golgins.[3]

A recent study, in contrast, associated TMEM87A with mechanosensitive cation channel activity in cultured melanoma cells.[4] Deflection of micropillars on which these cells were cultured resulted in cationic currents that were reduced by TMEM87A expression knockdown. Additionally, TMEM87A over-expression was sufficient to generate similar currents in PIEZO1-KO HEK293T cells.[4] When TMEM87A was knocked-out of the cultured melanoma cells, these cells also demonstrated an increase in cell adhesion strength and a decrease in cell migration, suggesting potential roles for TMEM87A in these processes in cancers.[4]

TMEM87B has been linked to developmental processes through a distinct line of inquiry in relation to recurrent 2q13 microdeletion syndrome.[5,8-9] Patients with this syndrome exhibit cardiac defects, craniofacial anomalies, and developmental delay. The vast majority of these

microdeletions include TMEM87B.[8] In zebrafish, morpholino knockdown of TMEM87B resulted in cardiac hypoplasia.[5] Finally, whole-exome sequencing uncovered a paternally inherited single missense mutation in TMEM87B in a 2q13 microdeletion syndrome patient with a severe cardiac phenotype; the deletion at chromosome 2 was maternally inherited from an unaffected mother.[8] Both TMEM87A and TMEM87B have additionally been implicated in cancers such as non-small cell lung cancer through fusion or suspected interaction with oncogenes.[6-7,10]

Bioinformatic analysis has grouped TMEM87A and TMEM87B in a small family of proteins (termed lung 7TM receptors or LUSTRs), which also includes the orphan GPCRs GPR107 and GPR108.[1] Notably, GPR107 and GPR108 similarly localize to the Golgi and have suggested roles in protein transport. GPR107 is implicated in the transport of *P. aeruginosa* Exotoxin A, *C. jejuni* cytolethal distending toxin (CDT), and ricin and GPR107-KO mice display deficits in receptor mediated endocytosis and recycling.[11-14] GPR108 is critical for the transduction of a majority of AAV serotypes.[15-16] While the GPR107-KO mouse was found to be embryonic lethal, the GPR108-KO mouse is viable and uncovered a potential role for GPR108 in the regulation of Toll-like receptor (TLR) mediated signaling.[14,17]

Despite being implicated in important cell biological processes and linked to human health, our molecular understanding of TMEM87s or of any LUSTR family proteins is limited. This is part because there is no experimental structural information available for these proteins to date. Here, we present a cryo-electron microscopy (cryo-EM) structure of human TMEM87A. Through structural and bioinformatic analyses, we find TMEM87A belongs to a family of proteins with a characteristic domain organization and structural homology we term GOLD-domain seven transmembrane (GOST) proteins. We speculate TMEM87A and other GOST proteins could function as trafficking chaperones for hydrophobic cargo.

4.3 Results

4.3.1 Structure determination

Full-length human TMEM87A was expressed and purified from Sf9 cells and reconstituted into lipid nanodiscs composed of MSP1D1 and a mixture of DOPE, POPS, and POPC lipids (Supplemental Figure 4.1). Cryo-electron microscopy (cryo-EM) was used to determine the structure of TMEM87A to a nominal resolution of 4.7 Å, with better resolved regions reaching ~4.3 Å in the core of the protein. The map was of sufficient quality to place secondary structure elements unambiguously and, using an AlphaFold predicted structure as a starting model, to build a model consisting of 395/555 residues (Fig. 4.1, Supplemental Figure 4.2).[18] The modeled structure and AlphaFold predicted structure are similar (overall r.m.s.d. = 2.0 Å) with one notable difference being the relative position of extracellular and transmembrane regions. 37 N-terminal residues (amino acids 1-37), 88 C-terminal residues (475-555), and 36 residues within loops in the extracellular domain (145-173, 192-198) were not resolved in the density and are unmodeled.

TMEM87A is composed of an extracellular beta-sandwich domain atop a G-protein coupled receptor (GPCR)-like seven transmembrane domain (Fig. 4.1). The extracellular beta-sandwich

domain is formed by seven beta strands in two opposing sheets, with three strands on the N-terminal face and four on the opposing face. The N terminal face consists of β1, β3, β7, and a region between β5 and β6 that lacks secondary structure and packs against β3 at the top of the sheet. The opposing face of the β-sandwich is formed from β2, β4, β5, β6 with a helix-turn-helix motif between β4 and β5. A short 10 residue loop extends from the end of β7 to form the start of H1. Two disulfide bonds are observed. The first, between C74 and C128, connects two loops at the top of this extracellular domain. The second, between C89 and C431, tethers the bottom of the extracellular domain to the top of the transmembrane domain, perhaps constraining their relative movement.

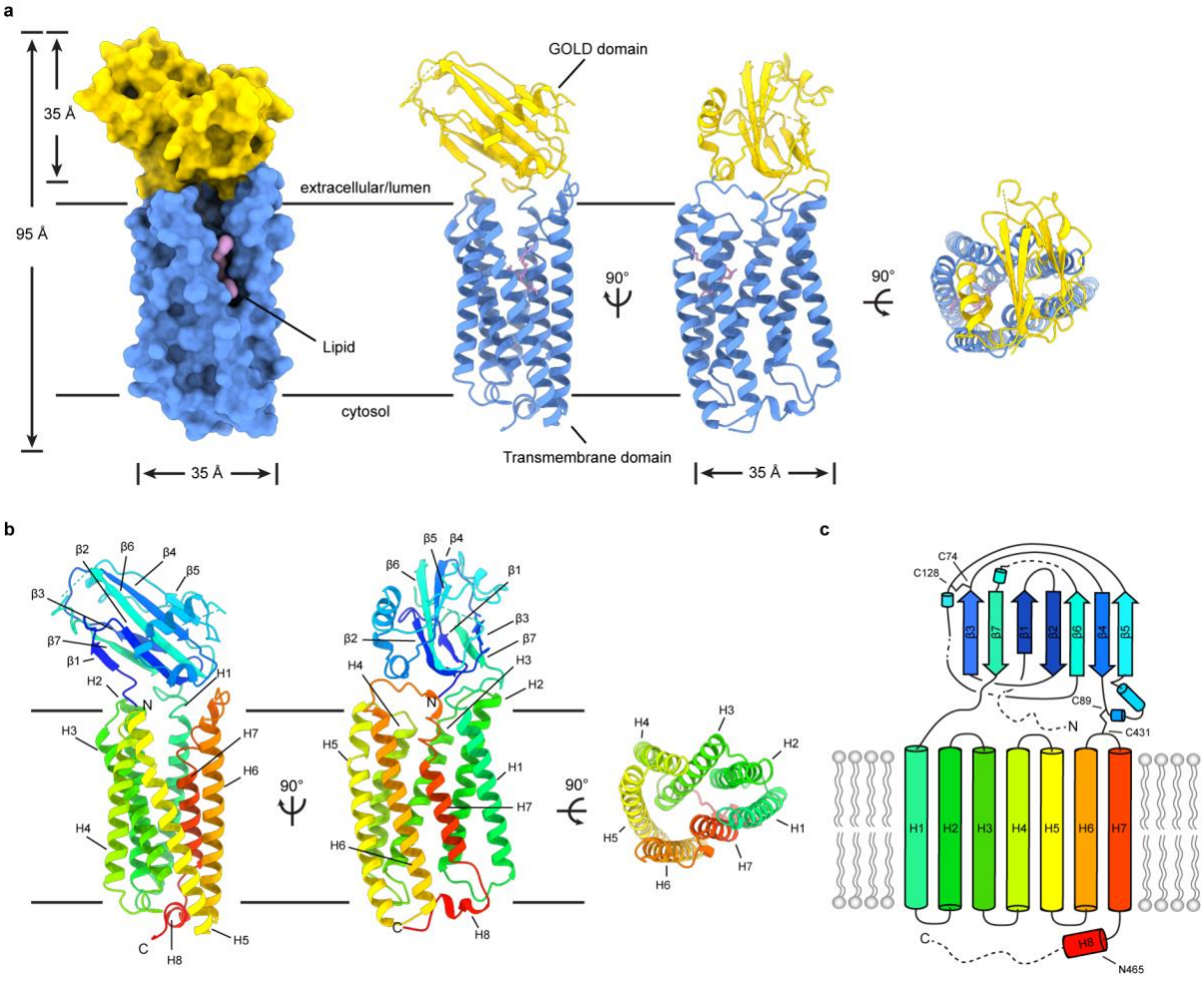


Figure 4.1: Architecture of TMEM87A in lipid nanodiscs. a, Model for TMEM87A viewed from the plane of the membrane. Extracellular GOLD domain is colored yellow, seven-transmembrane domain colored blue, and phospholipid colored pink. **b,** The TMEM87A protomer with rainbow coloring from N-terminus (blue) to C-terminus (red). **c,** Corresponding cartoon of TMEM87A domain topology with rainbow coloring from N-terminus (blue) to C-terminus (red). Disulfide bonds and residues mentioned throughout the text noted.

4.3.2 Similarity to seven-transmembrane proteins

The transmembrane region of TMEM87A is generally structurally similar to other seven-transmembrane proteins (7TM), with a few notable differences (Supplemental Figure 4.3). Using

Dali to compare the isolated TMEM87A transmembrane domain to all experimentally determined protein structures returns 7TM proteins including the fungal class D GPCR Ste2, microbial opsins, and additional class A GPCRs as clear structural homologs.[19-22] Since TMEM87A shares similar structural features with ion conducting opsins and previous work implicated TMEM87A in mechanosensitive cation conduction,[4] we asked whether TMEM87A displayed channel activity in isolation. We purified and reconstituted TMEM87A into proteoliposomes and recorded currents across patched membranes in response to membrane stretch induced by pressure steps. We observed neither basal or mechanically activated currents in TMEM87A reconstituted proteoliposomes, in contrast to the mechanosensitive ion channel TRAAK used as a positive control (Supplemental Figure 4.4). This result is consistent with the lack of a clear ion conducting path in the TMEM87A structure. We conclude that under these limited reductionist conditions TMEM87A does not form a mechanosensitive ion channel. We can not fully rule out the possibility that TMEM87A contributes to mechanosensitive ion channel currents under other conditions, in complex with other proteins, or in response to additional factors. Still, we explored whether alternative functions are consistent with TMEM87A structure.

We next asked whether TMEM87A shared structural features consistent with G-protein coupled receptor activity. TMEM87A and GPCRs have a similar transmembrane helix organization, for example, TMEM87A and the β1-adrenergic receptor are superimposed with a r.m.s.d. of 4.13 Å (Fig. 4.2).[23] Intriguingly, the position of cytoplasmic helix 8 in TMEM87A differs substantially from experimental structures of GPCRs. In TMEM87A, helix 8 turns back towards the center of the protein and packs against the bottom of the transmembrane domain (Fig. 4.2a,b). In experimental GPCR structures, helix 8 is rotated nearly 180º to adopt a different position and rather projects towards H1 and the surrounding membrane (Fig. 4.2b-c).[23-25] The position of H8 is necessary to accommodate G-protein or arrestin binding as seen in structures of the β1-adrenergic receptor (and is similarly positioned in isolated and complexed receptor structures). The position of H8 in TMEM87A would sterically clash with putative G-protein or arrestin binding through interfaces analogous to those observed in GPCR structures (Fig. 4.2c-e). Notably, the TMEM87B point mutation implicated in 2q13 deletion syndrome (TMEM87B N456D) corresponds to TMEM87A residue N465 at the H7-helix 8 junction (Fig. 4.2a)[8], suggesting helix 8 is important for its function. In addition to the difference in H8 position, TMEM87A does not share canonical motifs of class A or D GPCRs including the class A NPxxY activation motif, PIF motif, or D(E)/RY motif and class D LPLSSMWA activation motif.[20,26-27] Taken together, these differences suggest TMEM87A may not couple to G proteins or arrestins in a manner analogous to canonical GPCRs.

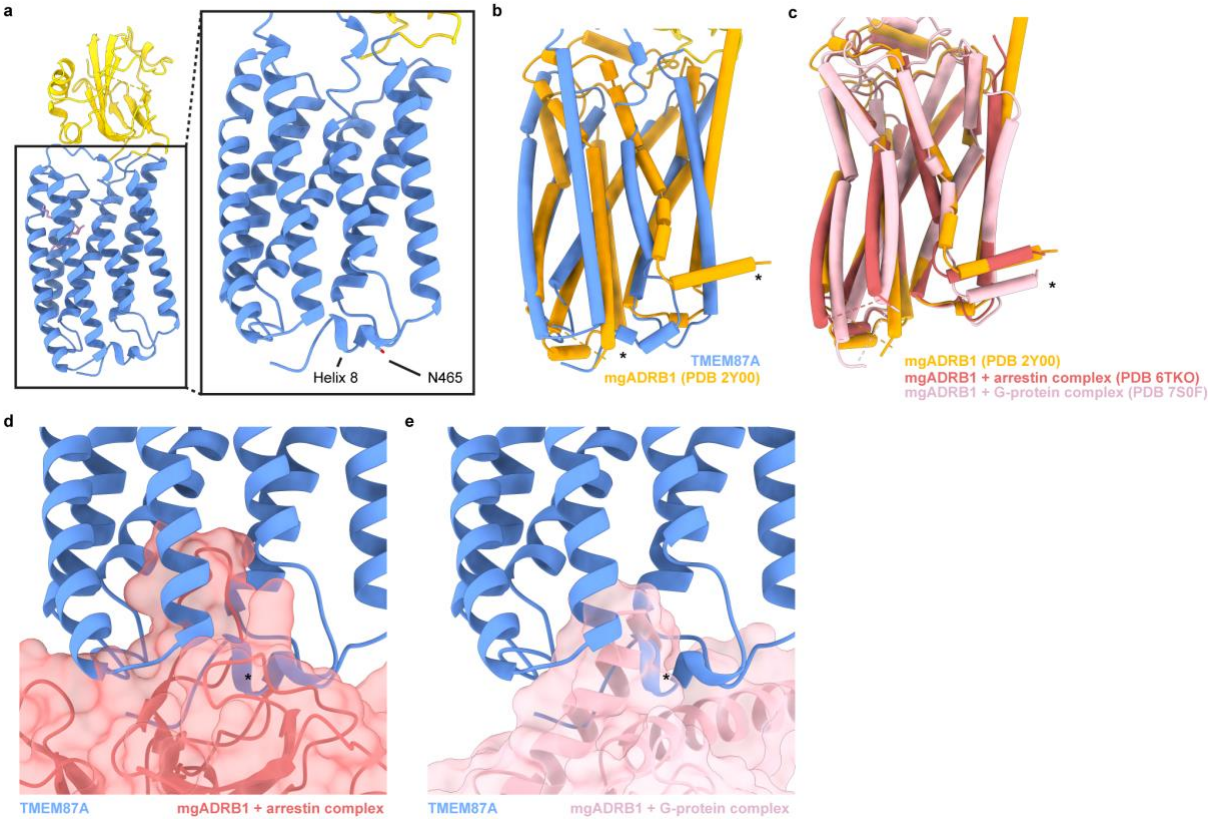


Figure 4.2: TMEM87A transmembrane domain and helix 8. **a,** View of the TMEM87A model from the "back" in the plane of the membrane and zoomed view highlighting the position of helix 8. Residue N465 on helix 8 is shown. **b,** Overlay of TMEM87A in blue and mgADRB1 (PDB 2Y00) in orange transmembrane domains, shown as tubes. **c,** Overlay of the transmembrane domains, shown as tubes, of mgADRB1 (PDB 2Y00) in orange with mgADRB1 + arrestin complex (PDB 6TKO) in red and mgADRB1 + G-protein complex (PDB 7S0F) in pink. **d,** Overlay of TMEM87A in blue and arrestin from the mgADRB1 complex (PDB 6TKO) in red transparent surface. **e,** Overlay of TMEM87A in blue and G-protein from the mgADRB1 complex (PDB 7S0F) in pink transparent surface. Asterisks denote the positions of helix 8 throughout.

4.3.3 Identification of TMEM87A GOLD domain and GOST protein family

We found the extracellular domain of TMEM87A adopts a Golgi dynamics (GOLD) domain fold. Using Dali to compare the isolated TMEM87A extracellular β-sandwich domain (residues 38-218) to all experimentally determined protein structures returns human p24 proteins as the top hits. Superposition of the TMEM87A extracellular domain and p24 shows conservation of beta strand topology throughout the domain, with differences and conservation of indeed contains an N terminal GOLD domain based on clear homology between the structures (Supplemental Figure 4.5a-c).[19,28] GOLD domain proteins, including p24 proteins, have a clear established role in the secretory pathway.[28,29] p24 proteins in particular are believed to act as cargo receptors for protein transport, with the GOLD domain mediating cargo recognition.[30-31] When TMEM87A's GOLD domain is compared to the GOLD domain of p24beta1 we see that both are β-sandwiches with similar topology, yet p24beta1 lacks the helix-turn-helix motif observed in TMEM87A's GOLD domain (overall r.m.s.d 2.62 Å) (Supplemental Figure 4.5c). Among p24 proteins the GOLD domain loops are one of the most variable regions and have been hypothesized to differentially recognize target cargo.[28] We see clear differences in the loops between these two GOLD domain

proteins, perhaps suggesting differences in interacting partners. This identification of TMEM87A's GOLD domain provides a potential structural explanation for the role of TMEM87 proteins in protein transport.

We additionally performed a Dali search against all predicted structures from the human proteome in the AlphaFold Protein Structure Database, using our entire experimental structure as the reference.[18-19,32] This search uncovered 7 additional proteins for which the predicted structure was an N terminal GOLD domain fused to a GPCR-like seven transmembrane domain. Notably, this search identified four new proteins with this fold: GPR180, TMEM145, TMEM181, and WLS, in addition to identifying all four members of the LUSTR family (Fig. 4.3, Supplemental Figure 4.5d-j, Supplemental Figure 4.6a-g). GPR180 has been implicated in TGF-B signaling, TMEM181 has been found to play a key role in E. coli cytolethal distending toxin toxicity, and TMEM145 has been minimally characterized to date.[33-34] Intriguingly, WLS has been very well characterized for its role in in the transport of Wnts and the KO of WLS in mice is embryonic lethal.[35-37] For all 7 of these proteins, the predicted structure closely resembles the structure of TMEM87A (overall r.m.s.d range from 1.85 to 3.87 Å). We propose the expansion of the LUSTR family of proteins to include these additional members, and to name this new family: GOLD domain seven transmembrane (GOST) proteins.

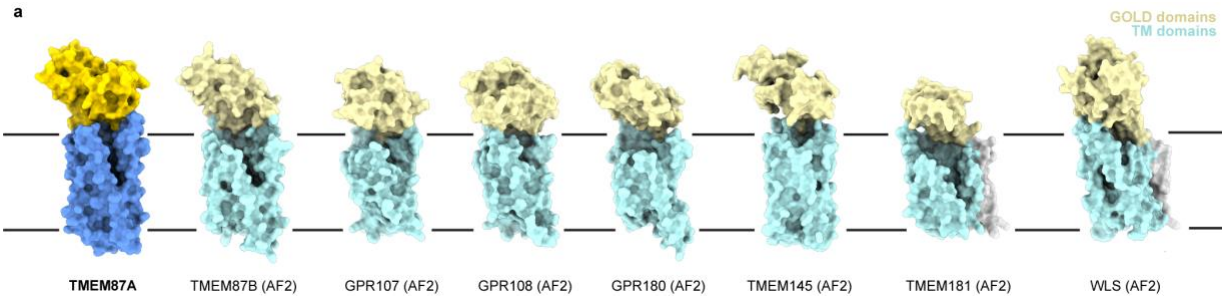


Figure 4.3: AlphaFold2 predictions of GOST family protein structures. a, Comparison of present TMEM87A structure with AlphaFold2 predictions of other GOST family proteins: TMEM87B, GPR107, GPR108, GPR180, TMEM145, TMEM181, and WLS. Predicted structures have gold domains colored light yellow, seven-transmembrane domain colored light blue, and if applicable, N terminal helix colored light gray. Predicted structures are only displaying regions of the structure with a confidence score of >=80 pLDDT.

4.3.4 Structural comparisons of TMEM87A and WLS

Recent cryo-EM structures of WLS bound to Wnt provide an example of a GOST protein bound to its corresponding cargo (Fig. 4.4a).[38-39] We find that the transmembrane region of TMEM87A aligns well with WLS with one notable rearrangement, H4 and H5 in WLS are shifted approximately 13 Å away from the center of the protein, resulting in a larger cavity, likely necessary to accommodate the insertion of the lipidated domain of Wnt8a (Fig. 4.4b). We also observe a rearrangement of the GOLD domain, with the WLS GOLD domain shifted approximately 26 Å in the same direction of movement as the H4/H5 shift (Fig. 4.4b). Intriguingly, TMEM87A and WLS exhibit similar overall patterns of surface electrostatic charge. In WLS the binding pocket for the Wnt palmitoyl group is an electronegative surface, and we observe a similar electronegative pocket in the cavity of TMEM87A (Fig. 4.4c-d).

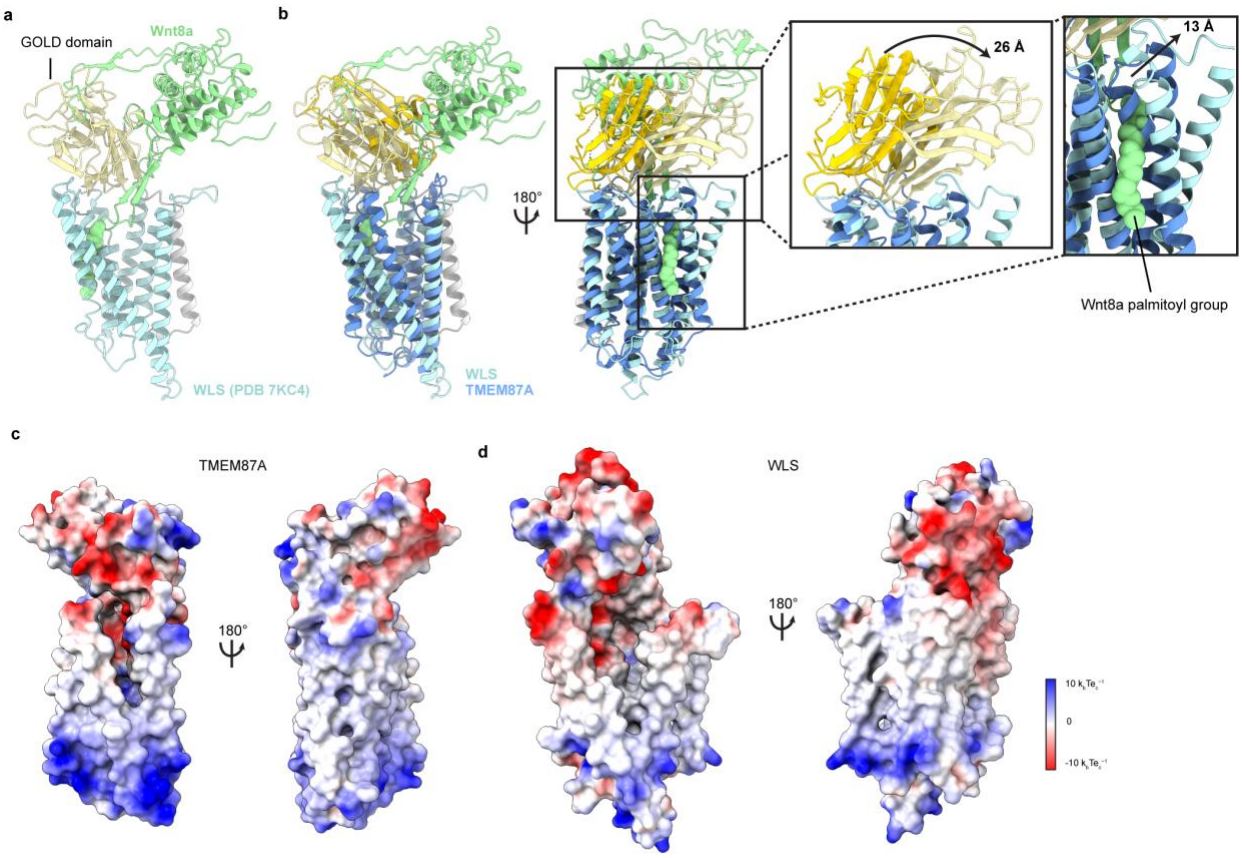


Figure 4.4: Structural comparison of TMEM87A and WLS bound to Wnt. a, Structure (PDB 7KC4) for WLS bound to Wnt8a with the WLS GOLD domain colored light yellow, the WLS seven-transmembrane domain colored light blue, the N terminal helix colored light gray, and Wnt8a colored light green. **b,** Overlay of the experimental structures of WLS (PDB 7KC4) and TMEM87A, with TMEM87A GOLD domain colored yellow and TMEM87A transmembrane domain colored blue. Zoomed in views highlight the structural rearrangements of WLS relative to TMEM87A. **c,** Electrostatic surface of TMEM87A. **d,** Electrostatic surface of WLS. Electrostatic potential scale is from -10 $k_bTe_c^{-1}$ to 10 $k_bTe_c^{-1}$ as indicated in key.

One of the most striking features of the TMEM87As structure, also observed in the experimental WLS structure, is the presence of a large cavity within the transmembrane domain (Fig. 4.5a-b). In TMEM87A this cavity measures approximately 8637 $Å^3$. This cavity has a large opening on the extracellular side and travels most of the length of the membrane, before being blocked by the position of helix 8. In WLS this cavity is largely filled by the insertion of the palmitoylated hairpin 2 from Wnt8a (Fig. 4.5c). Interestingly, when the experimental and predicted WLS structures are compared, it appears that WLS undergoes a conformational change resulting in a larger cavity, likely to accommodate Wnt binding (Supplemental Figure 4.6h-j). In TMEM87A the position of H5 and H6 is such to result in a large "slot" which exposes this large inner cavity to the upper leaflet of the bilayer (Fig. 4.5d). This portion of the TMEM87A cavity is occupied by an observed phospholipid, and a similarly positioned phospholipid is observed in the WLS structure (Fig. 4.5d-e). TMEM87A and WLS also display a similar pattern when sequence conservation is mapped onto the protein surface, with the most strongly conserved residues found in the transmembrane domain, especially those that line the internal cavity, and much less conservation in the respective GOLD domains (Supplemental Figure 4.7). Taken together, these

similarities suggest that TMEM87A could perhaps be competent to bind lipidated cargo through a mechanism similar to WLS binding of Wnts.

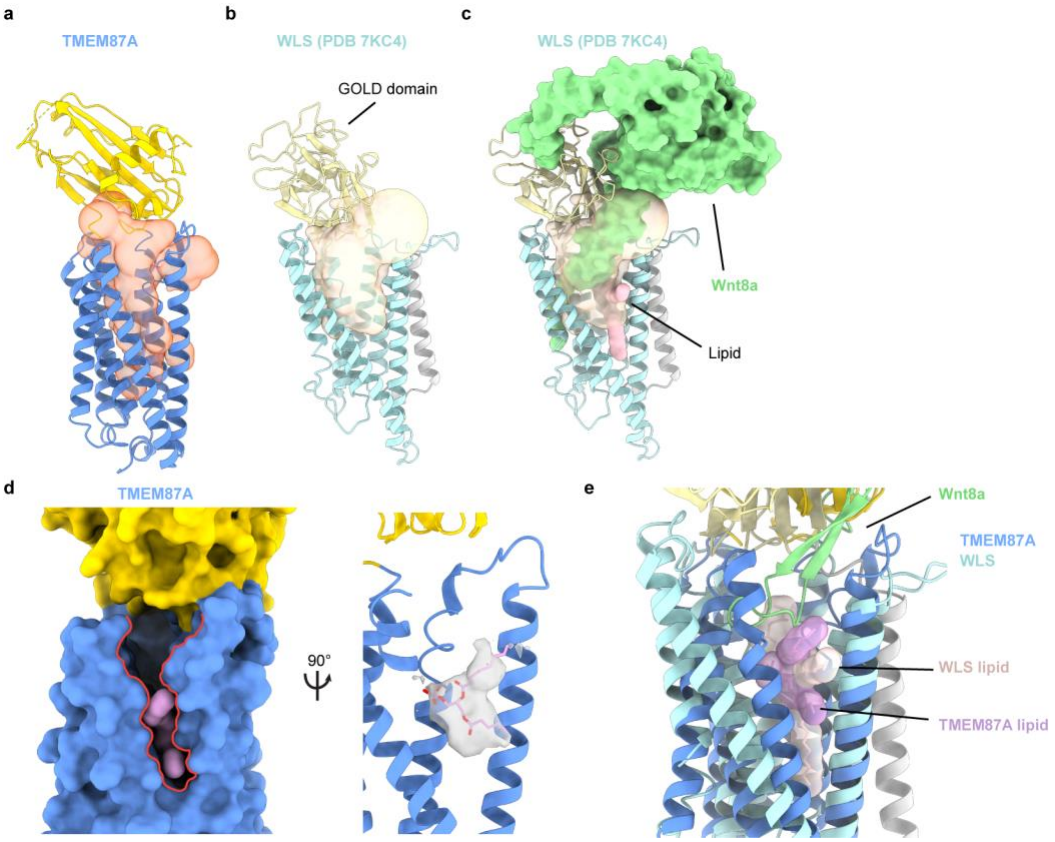


Figure 4.5: TMEM87A and WLS internal cavities. a, TMEM87A with CASTp calculated internal cavity colored in transparent orange surface. **b,** WLS (PDB 7KC4) with CASTp calculated internal cavity colored in transparent yellow surface. **c,** WLS cavity is filled by Wnt8a (shown as pale green surface) and phospholipid (pink). **d,** TMEM87A cavity has a "slot" exposed to the upper leaflet of the bilayer (outlined in red), with phospholipid observed in the lower region. Rotated view shows cryo-EM density corresponding to the modeled phospholipid. **e,** Overlay of the experimental TMEM87A and WLS structures showing the relative positions of phospholipids observed in each structure (colored hot pink transparent surface for TMEM87A, and light pink transparent surface for WLS).

4.4 Discussion

We have presented the first experimental structure of a TMEM87 protein, and through identification of its GOLD domain provided structural evidence to support a role in protein transport. In in vitro reconstituted proteoliposomes we find no evidence of mechanosensitive ion currents, nor does the structure make evident a likely ion permeation pathway. While these observations do not preclude that TMEM87A could act directly as an ion channel, perhaps TMEM87A acts on this previously described mechano-transduction pathway through the regulation of the transport or expression of a mechanosensitive ion channel.[4]

While the newly characterized GOST family of proteins remains largely understudied, a few defining elements have become clear. First, GOST family proteins have a clear role in protein

transport, as evidenced by both the presence of their respective GOLD domains, as well as prior evidence in the literature implicating 6/8 members in transport.[2-3,11-13,15-16,34-36] Second, a lack of clarity regarding GOST family proteins competency for GPCR signaling. While there have been some reports suggesting the identification of an endogenous ligand for some GOST proteins, GPR180 has been reported to not signal as a GPCR, and there is not any clear molecular evidence of GPCR signaling by a GOST protein.[40-41] Given the position of helix 8 in the present structure of TMEM87A and steric incompatibility with the binding of arrestins or G-proteins, this raises the question if GOST proteins in general are competent to act through GPCR signaling. Finally, the inclusion of WLS as a well-characterized member of the GOST family provides a potential model for understanding how these proteins might function.

These comparisons between experimental structures of TMEM87A and WLS, alongside predicted structures for the GOST family present the intriguing possibility that GOST family proteins could act specifically to transport lipidated secreted proteins through a mechanism analogous to Wnt binding and transport by WLS, potentially including the binding and transport of Wnts by other GOST family members. In such a mechanism, the two domains of the GOST family perform a sort of "coincidence detection" in which the GOLD domain is responsible for the identification and binding of the target cargo, and the internal cavity makes additional interactions including with lipidated or hydrophobic aspects of the cargo. GOLD domain proteins such as p24 have been implicated transport Wnts previously, and a number of important secreted proteins such as ghrelin and cytokines are known to be lipidated.[42-45] Future studies of GOST family proteins will be important to determine if the family could be broadly responsible for the transport of such lipidated and secreted proteins.

4.5 Methods

4.5.1 Cloning, expression and purification

The coding sequence for Homo sapiens TMEM87A (Uniprot ID: Q8NBN3) was codon-optimized for expression in Spodoptera frugiperda cells (Sf9 cells) and synthesized (Integrated DNA Technologies). The sequence was cloned into a custom vector based on the pACEBAC1 backbone (MultiBac, Geneva Biotech) with an added C-terminal PreScission protease (PPX) cleavage site, linker sequence, superfolder GFP (sfGFP) and 7×His tag, generating a construct for expression of TMEM87A-SNS-LEVLFQGP-SRGGSGAAAGSGSGS-sfGFP-GSS-7×His. MultiBac cells were used to generate a Bacmid according to the manufacturer's instructions. Sf9 cells were cultured in ESF 921 medium (Expression Systems) and P1 virus was generated from cells transfected with FuGENE transfection reagent (Active Motif) according to the manufacturer's instructions. P2 virus was then generated by infecting cells at 2 million cells per ml with P1 virus at a multiplicity of infection of roughly 0.1, with infection monitored by fluorescence and harvested at 72 h. P3 virus was generated in a similar manner to expand the viral stock. The P3 viral stock was then used to infect Sf9 cells at 2 million cells per ml at a multiplicity of infection of around 2~5. At 72 h, infected cells containing expressed TMEM87A-sfGFP protein were collected by centrifugation at 1,000g for 10 min and frozen at −80 °C. A cell pellet from 1 L culture was thawed and lysed by sonication in 100 mL buffer containing 50 mM Tris pH 8.0, 150 mM NaCl, 1 mM EDTA and protease inhibitors (1 mM phenylmethysulfonyl fluoride, 1 μM E64, 1 μg/mL pepstatin A, 10 μg/mL soy trypsin inhibitor, 1 μM benzimidine,

1 μg/mL aprotinin, and 1 μg/mL leupeptin). The membrane fraction was collected by centrifugation at 150, 000g for 45 min and homogenized with a cold Dounce homogenizer in 100 mL buffer containing 20 mM Tris pH 8.0, 150 mM NaCl, 1 mM EDTA, 1% n-dodecyl-b-D-maltopyranoside (DDM), 0.2% cholesteryl hemisuccinate (CHS) and protease inhibitors. Protein was extracted with gentle stirring for 2 h at 4 ℃. The extraction mixture was centrifuged at 33,000g for 45 min and the supernatant was bound to 5 mL Sepharose resin coupled to anti-GFP nanobody for 2 h at 4 ℃. The resin was the collected in a column and washed with 25 mL buffer 1 (20 mM Tris pH 8.0, 150 mM NaCl, 1 mM EDTA, 0.025% DDM, 0.005% CHS), 50 mL buffer 2 (20 mM Tris pH 8.0, 500 mM NaCl, 1 mM EDTA, 0.025% DDM, 0.005% CHS) and 25 mL buffer 1. The resin was then resuspended in 5 mL of buffer 1 with 0.5 mg PPX protease and rocked gently in the capped column overnight. Cleaved TMEM87A was eluted with an additional 8 mL buffer 1, spin concentrated to roughly 500 μL with Amicon Ultra spin concentrator 50-kDa cutoff (Millipore), and then loaded onto a Superose 6 increase column (GE Healthcare) on an NGC system (Bio-Rad) equilibrated in buffer 1. Peak fractions containing TMEM87A were then collected and spin concentrated before incorporation into proteoliposomes or nanodiscs.

4.5.2 Nanodisc reconstitution

Freshly purified TMEM87A was reconstituted into MSP1D1 nanodiscs with a mixture of lipids (DOPE:POPS:POPC at a 2:1:1 mass ratio, Avanti) at a final molar ratio of 1:4:400 (TMEM87A:MSP1D1:lipid mixture).[46] First, 20 mM solubilized lipid in nanodisc formation buffer (20 mM Tris pH8.0, 150 mM NaCl, 1 mM EDTA) was mixed with additional DDM detergent and TMEM87A. This solution was mixed at 4 °C for 30 min before addition of purified MSP1D1. The solution with MSP1D1 was mixed at 4 °C for 30 min before addition of 200 mg of Biobeads SM2. This mix was incubated at 4 °C for 30 min before addition of another 200 mg of Biobeads. This final mixture was then gently tumbled at 4 °C overnight (roughly 12 h). Supernatant was cleared of beads by letting large beads settle and carefully removing liquid with a pipette. Sample was spun for 10 min at 21,000g before loading onto a Superose 6 increase column in buffer containing 20 mM Tris pH8.0, 150 mM NaCl. Peak fractions corresponding to TMEM87A in MSP1D1 were collected, 50-kDa cutoff spin concentrated and used for grid preparation. MSP1D1 was prepared as described without cleavage of the His-tag.

4.5.3 EM sample preparation and data collection

TMEM87A in MSP1D1 nanodiscs was centrifuged at 21,000 × g for 5 min at 4 °C. A 3 μL sample was applied to holey carbon, 300 mesh R1.2/1.3 gold grids (Quantifoil, Großlöbichau, Germany) that were freshly glow discharged for 25 s. Sample was incubated for 5 s at 4 °C and 100% humidity prior to blotting with Whatman #1 filter paper for 3 s at blot force 1 and plunge-freezing in liquid ethane cooled by liquid nitrogen using a FEI Mark IV Vitrobot (FEI/Thermo Scientific, USA). Grids were clipped and transferred to a FEI Talos Arctica electron microscope operated at 200 kV. Fifty frame movies were recorded on a Gatan K3 Summit direct electron detector in super-resolution counting mode with pixel size of 0.5575 Å. The electron dose rate was 8.849 e− Å2 s−1 and the total dose was 50.0 e− Å2. Nine movies were collected around a central hole position with image shift and defocus was varied from −0.6 to −1.8 μm through SerialEM.[47]

See Table 4.1 for data collection statistics.

4.5.4 Cryo-EM data processing

Motion-correction with dose-weighting was performed using RELION3.1's implementation of MotionCor2 and "binned" 2x from super-resolution to the physical pixel size.[48-50] CTFFIND-4.1 was then used to estimate the contrast transfer function (CTF).[51] Micrographs with a CTF maximum estimated resolution lower than 5 Å were discarded. Particle images were picked first with RELION3.1's Laplacian-of-Gaussian filter, then following initial clean-up and 2D-classification, templated-based auto-picking was performed. For this particle set, 2D-classification was iteratively performed in both RELION3.1 and cryoSPARC v2, then iterative ab initio and heterogeneous refinement jobs were used to further identify higher quality particles.[52] Once an initial high quality set of particles was determined, these particle positions were used for training in Topaz, and the resulting Topaz model used to repick particles.[53] The above pipeline was applied again to the Topaz picked particles and the resulting set of particles was merged with the initial set of template-picked particles, and duplicates removed. This final set of particles was then input to Bayesian particle polishing in RELION3.1. The output "shiny" particles were then input to iterative homogenous and nonuniform refinements in cryoSPARC v2 until no further improvements were observed.[54] The output of the best nonuniform refinement was used for particle subtraction in RELION3.1, after which final refinement and postprocessing jobs were performed. This final refinement was input to Phenix resolve density modification to generate an additional map used during modeling.[55] The initial resolution and dynamic mask nonuniform parameters were adjusted empirically to yield the best performing refinement. UCSF pyem was used for conversion of files from cryoSPARC to Relion formats.[56]

4.5.5 Model building and refinement

The final relion postprocessed map and Phenix density modified map were used for modeling. The AlphaFold2 TMEM87A model was rigid body fit to the density in ChimeraX and used as a foundation for manual model building in Coot.[18,32,57-58] The model was real space refined in Phenix and assessed for proper stereochemistry and geometry using Molprobity.[59-60] FSCs calculated in Phenix mtriage. Structures were analyzed and figures were prepared with CASTp, DALI, ChimeraX, JalView, Prism 8, Python GNU Image Manipulation Program, and Adobe Photoshop and Illustrator software.[19,61-62] Consurf was used to map conservation onto the structure surfaces using an alignment of sequences determined using SHOOT.[63-64] Electrostatic potentials calculated in ChimeraX. CASTp cavity calculation output was converted from JSON to PDB using a custom python script.

4.5.6 Proteoliposome reconstitution

For proteoliposome patching experiments, we incorporated protein into lipid and generated proteoliposome blisters for patch recordings using dehydration and rehydration as described previously with the following modifications. TMEM87A in buffer 1 was exchanged into soybean l-α-phosphatidylcholine (Soy PC, MillaporeSigma) with the addition of Biobeads SM2 (Bio-Rad) and an hour incubation at a protein:lipid ratio of 1:10 (corresponding to 0.4 mg purified TMEM87A and 4 mg of Soy PC lipid or 1:50 in buffer (5 mM HEPES pH 7.2, 200 mM KCl).

TRAAK control proteoliposomes were prepared at 1:50 as described previously. Control liposomes were prepared from the same lipid and protocol with protein replaced with buffer 1.

4.5.7 Electrophysiology

Proteoliposomes were thawed and dispensed in 0.5-1 μL drops on a 35 mm glass-bottom dish. The drops were dried in a vacuum chamber in the dark overnight. Proteoliposomes were rehydrated with 20 μL buffer (5 mM HEPES pH 7.2, 200 mM KCl). Each drop was firstly covered with a buffer and then let surface tension connect drops. Rehydrating proteoliposomes were placed within a humid chamber at 4 ℃ before patching. Recordings were made at room temperature using Clampex v.10.7 data acquisition software (as part of the pClamp v.10.7 suite) with an Axopatch 200B Patch Clamp amplifier and Digidata 1550B digitizer (Molecular Devices) at a bandwidth of 1 kHz and digitized at 500 kHz. A pressure clamp (ALA Scientific) was used to form seals. Pipette solution was 10 mM HEPES pH 7.2, 150 mM KCl, 3 mM MgCl2 and 5 mM EGTA. Bath solution was 10 mM HEPES pH 7.3, 135 mM NaCl, 15 mM KCl, 1 mM CaCl2, 3 mM MgCl2. Borosilicate glass pipettes were pulled and polished to a resistance of 2–5 MΩ when filled with pipette solution.

4.6 Data availability

All data associated with this study will be publicly available. The TMEM87A model will be deposited in the Protein Data Bank, the final cryo-EM maps will be in the Electron Microscopy Data Bank (EMDB), and the original micrograph movies and final particle stack will be the Electron Microscopy Public Image Archive (EMPIAR).

4.7 Acknowledgements

The authors thank J. Remis, D. Toso, and P. Tobias for microscope and computational support at the Cal-Cryo facility of UC Berkeley. The authors thank all members of the Brohawn lab for helpful discussions and critical feedback on the project. S.G.B. is a New York Stem Cell Foundation-Robertson Neuroscience Investigator. This work was funded by the New York Stem Cell Foundation; NIGMS grant no. GM123496; a McKnight Foundation Scholar Award; a Sloan Research Fellowship; and a Winkler Family Scholar Award (to S.G.B.)

4.8 Contributions

L.Z. generated expression constructs, purified proteins, prepared samples for cryo-EM and collected cryo-EM data. C.M.H and L.Z. processed cryo-EM data. C.M.H. performed structural analysis. C.M.H and L.Z. performed electrophysiology experiments. C.M.H. and S.G.B. modeled the structure. C.M.H and S.G.B. wrote the manuscript with input from L.Z. S.G.B supervised the project.

4.9 References

1. Edgar, Alasdair J. Human GPR107 and murine Gpr108 are members of the LUSTR family of proteins found in both plants and animals, having similar topology to G-protein coupled receptors. *DNA Sequence* **18**, 235–241 (2007).
2. Hirata, T. *et al.* Post-Golgi anterograde transport requires GARP-dependent endosome-to-TGN retrograde transport. *MBoC* **26**, 3071–3084 (2015).
3. Shin, J. J. H. *et al.* Spatial proteomics defines the content of trafficking vesicles captured by golgin tethers. *Nat Commun* **11**, 5987 (2020).
4. Patkunarajah, A. *et al.* TMEM87a/Elkin1, a component of a novel mechanoelectrical transduction pathway, modulates melanoma adhesion and migration. *eLife* **9**, e53308 (2020).
5. Russell, M. W. *et al.* Functional analysis of candidate genes in 2q13 deletion syndrome implicates FBLN7 and TMEM87B deficiency in congenital heart defects and FBLN7 in craniofacial malformations. *Human Molecular Genetics* **23**, 4272–4284 (2014).
6. Cooper, A. J. *et al.* Identification of a RAS-activating *TMEM87A–RASGRF1* Fusion in an Exceptional Responder to Sunitinib with Non–Small Cell Lung Cancer. *Clin Cancer Res* **26**, 4072–4079 (2020).
7. Shaver, T. M. *et al.* Diverse, Biologically Relevant, and Targetable Gene Rearrangements in Triple-Negative Breast Cancer and Other Malignancies. *Cancer Res* **76**, 4850–4860 (2016).
8. Yu, H.-C. *et al.* Discovery of a potentially deleterious variant in *TMEM87B* in a patient with a hemizygous 2q13 microdeletion suggests a recessive condition characterized by congenital heart disease and restrictive cardiomyopathy. *Cold Spring Harb Mol Case Stud* **2**, a000844 (2016).
9. Digilio, M. C. *et al.* Congenital heart defects in the recurrent 2q13 deletion syndrome. *European Journal of Medical Genetics* **65**, 104381 (2022).
10. Li, S. *et al.* Expression of TMEM87B interacting with the human papillomavirus type 18 E6 oncogene in the Hela cDNA library by a yeast two-hybrid system. *Oncol Rep* (2008).
11. Tafesse, F. G. *et al.* GPR107, a G-protein-coupled Receptor Essential for Intoxication by Pseudomonas aeruginosa Exotoxin A, Localizes to the Golgi and Is Cleaved by Furin. *Journal of Biological Chemistry* **289**, 24005–24018 (2014).
12. Carette, J. E. *et al.* Global gene disruption in human cells to assign genes to phenotypes by deep sequencing. *Nat Biotechnol* **29**, 542–546 (2011).
13. Elling, U. *et al.* Forward and Reverse Genetics through Derivation of Haploid Mouse Embryonic Stem Cells. *Cell Stem Cell* **9**, 563–574 (2011).
14. Zhou, G. L., Na, S.-Y., Niedra, R. & Seed, B. Deficits in receptor-mediated endocytosis and recycling in cells from mice bearing a disruption of the *Gpr107* locus. *Journal of Cell Science* jcs.135269 (2014)
15. Dudek, A. M. *et al.* GPR108 Is a Highly Conserved AAV Entry Factor. *Molecular Therapy* **28**, 367–381 (2020).
16. Meisen, W. H. *et al.* Pooled Screens Identify GPR108 and TM9SF2 as Host Cell Factors Critical for AAV Transduction. *Molecular Therapy - Methods & Clinical Development* **17**, 601–611 (2020).
17. Dong, D. *et al.* GPR108, an NF-κB activator suppressed by TIRAP, negatively regulates TLR-triggered immune responses. *PLoS ONE* **13**, e0205303 (2018).

18. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
19. Holm L (2020) Using Dali for protein structure comparison. Methods Mol. Biol. 2112, 29-42
20. Velazhahan, V. *et al.* Structure of the class D GPCR Ste2 dimer coupled to two G proteins. *Nature* **589**, 148–153 (2021).
21. Yamamoto, M., Hayakawa, N., Murakami, M. & Kouyama, T. Crystal Structures of Different Substates of Bacteriorhodopsin's M Intermediate at Various pH Levels. *Journal of Molecular Biology* **393**, 559–573 (2009).
22. Hayashi, T. *et al.* How Does a Microbial Rhodopsin RxR Realize Its Exceptionally High Thermostability with the Proton-Pumping Function Being Retained? *J. Phys. Chem. B* **124**, 990–1000 (2020).
23. Warne, T. *et al.* The structural basis for agonist and partial agonist action on a β1-adrenergic receptor. *Nature* **469**, 241–244 (2011).
24. Lee, Y. *et al.* Molecular basis of β-arrestin coupling to formoterol-bound β1-adrenoceptor. *Nature* **583**, 862–866 (2020).
25. Alegre, K. O. *et al.* Structural basis and mechanism of activation of two different families of G proteins by the same GPCR. *Nat Struct Mol Biol* **28**, 936–944 (2021).
26. Palczewski, K. *et al.* Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor. **289**, 8 (2000).
27. Wang, J., Hua, T., & Liu, Z-J. Structural features of activated GPCR signaling complexes. *Current Opinion in Structural Biology* **8** (2020).
28. Nagae, M. *et al.* 3D Structure and Interaction of p24β and p24δ Golgi Dynamics Domains: Implication for p24 Complex Formation and Cargo Transport. *Journal of Molecular Biology* **428**, 4087–4099 (2016).
29. Anantharaman, V. & Aravind, L. The GOLD domain, a novel protein module involved in Golgi function and secretion. *Genome Biology* **3** (5) (2002).
30. Pastor-Cantizano, N., Montesinos, J. C., Bernat-Silvestre, C., Marcote, M. J. & Aniento, F. p24 family proteins: key players in the regulation of trafficking along the secretory pathway. *Protoplasma* **253**, 967–985 (2016).
31. Mendes, L. F. S. & Costa-Filho, A. J. A gold revision of the Golgi Dynamics (GOLD) domain structure and associated cell functionalities. *FEBS Letters* **596**, 973–990 (2022).
32. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* **50**, D439–D444 (2022).
33. Balazova, L. *et al.* GPR180 is a component of TGFβ signalling that promotes thermogenic adipocyte function and mediates the metabolic effects of the adipocyte-secreted factor CTHRC1. *Nat Commun* **12**, 7144 (2021).
34. Carette, J. E. *et al.* Haploid Genetic Screens in Human Cells Identify Host Factors Used by Pathogens. *Science* **326**, 1231–1235 (2009).
35. Bänziger, C. *et al.* Wntless, a Conserved Membrane Protein Dedicated to the Secretion of Wnt Proteins from Signaling Cells. *Cell* **125**, 509–522 (2006).
36. Bartscherer, K., Pelte, N., Ingelfinger, D. & Boutros, M. Secretion of Wnt Ligands Requires Evi, a Conserved Transmembrane Protein. *Cell* **125**, 523–533 (2006).

37. Fu, J., Jiang, M., Mirando, A. J., Yu, H.-M. I. & Hsu, W. Reciprocal regulation of Wnt and Gpr177/mouse Wntless is required for embryonic axis formation. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 18598–18603 (2009).
38. Nygaard, R. *et al.* Structural Basis of WLS/Evi-Mediated Wnt Transport and Secretion. *Cell* **184**, 194-206.e14 (2021).
39. Zhong, Q. *et al.* Cryo-EM structure of human Wntless in complex with Wnt3a. *Nat Commun* **12**, 4541 (2021).
40. Mosienko, V. *et al.* Putative Receptors Underpinning l-Lactate Signalling in Locus Coeruleus. *Neuroglia* **1**, 365–380 (2018).
41. Yosten, G. L. C., Redlinger, L. J. & Samson, W. K. Evidence for an interaction of neuronostatin with the orphan G protein-coupled receptor, GPR107. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **303**, R941–R949 (2012).
42. Port, F., Hausmann, G. & Basler, K. A genome-wide RNA interference screen uncovers two p24 proteins as regulators of Wingless secretion. *EMBO Rep* **12**, 1144–1152 (2011).
43. Hang, H. C. & Linder, M. E. Exploring Protein Lipidation with Chemical Biology. *Chem. Rev.* **111**, 6341–6358 (2011).
44. Jiang, H. *et al.* Protein Lipidation: Occurrence, Mechanisms, Biological Functions, and Enabling Technologies. *Chem. Rev.* **118**, 919–988 (2018).
45. Chen, B., Sun, Y., Niu, J., Jarugumilli, G. K. & Wu, X. Protein Lipidation in Cell Signaling and Diseases: Function, Regulation, and Therapeutic Opportunities. *Cell Chemical Biology* **25**, 817–831 (2018).
46. Ritchie, T. K. *et al.* Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs. *Methods in Enzymology* **464**, 211–231 (2009).
47. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *Journal of Structural Biology* **152**, 36–51 (2005).
48. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331–332 (2017).
49. Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5–17 (2019).
50. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
51. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology* **192**, 216–221 (2015).
52. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290–296 (2017).
53. Bepler, T. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods* **16**, 13 (2019).
54. Punjani, A., Zhang, H., Fleet, D.J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nature Methods* **17**, 13 (2020).
55. Terwilliger, T. C., Ludtke, S. J., Read, R. J., Adams, P. D. & Afonine, P. V. Improvement of cryo-EM maps by density modification. *Nat Methods* **17**, 923–927 (2020).

56. Asarnow, D., Palovcak, E., Cheng, Y. UCSF pyem v0.5. *Zenodo* https://doi.org/10.5281/zenodo.3576630 (2019).
57. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis: UCSF ChimeraX Visualization System. *Protein Science* **27**, 14–25 (2018).
58. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *Coot*. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501 (2010).
59. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Crystallogr D Struct Biol* **75**, 861–877 (2019).
60. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation: PROTEIN SCIENCE.ORG. *Protein Science* **27**, 293–315 (2018).
61. Tian, W., Chen, C., Lei, X., Zhao, J. & Liang, J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Research* **46**, W363-W367 (2018).
62. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
63. Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* **44**, W344–W350 (2016).
64. Emms, D. M. and Kelley, S. SHOOT: phylogenetic gene search and ortholog inference. *Genome Biology* **23 (**85) (2022).

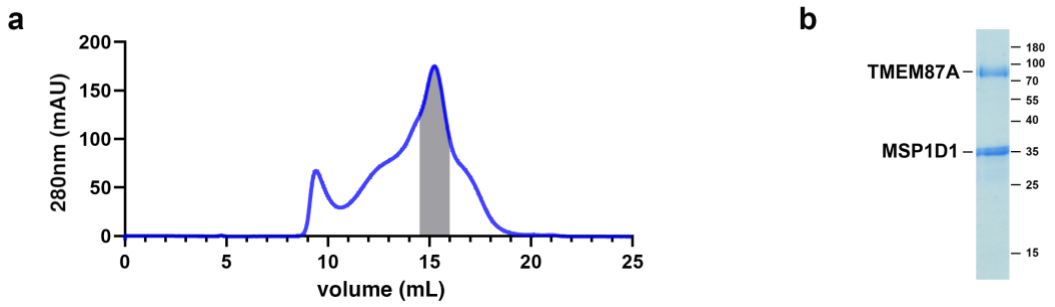
4.10 Tables

Data collection	**TMEM87A**
Total movies #	7060
Selected movies #	5898
Magnification	36,000 x
Voltage (KV)	200
Electron exposure ($e^-/Å^2$)	50
Frame #	50
Defocus range (um)	-0.6 to -1.8
Super resolution pixel size ($Å^2$)	0.5575
Binned pixel size ($Å^2$)	1.115
Processing	
Initial particle images (no.)	2,674,406
Final particle images (no.)	138,217
Map resolution Masked (Å, FSC = 0.143)	4.7
Symmetry imposed	C1

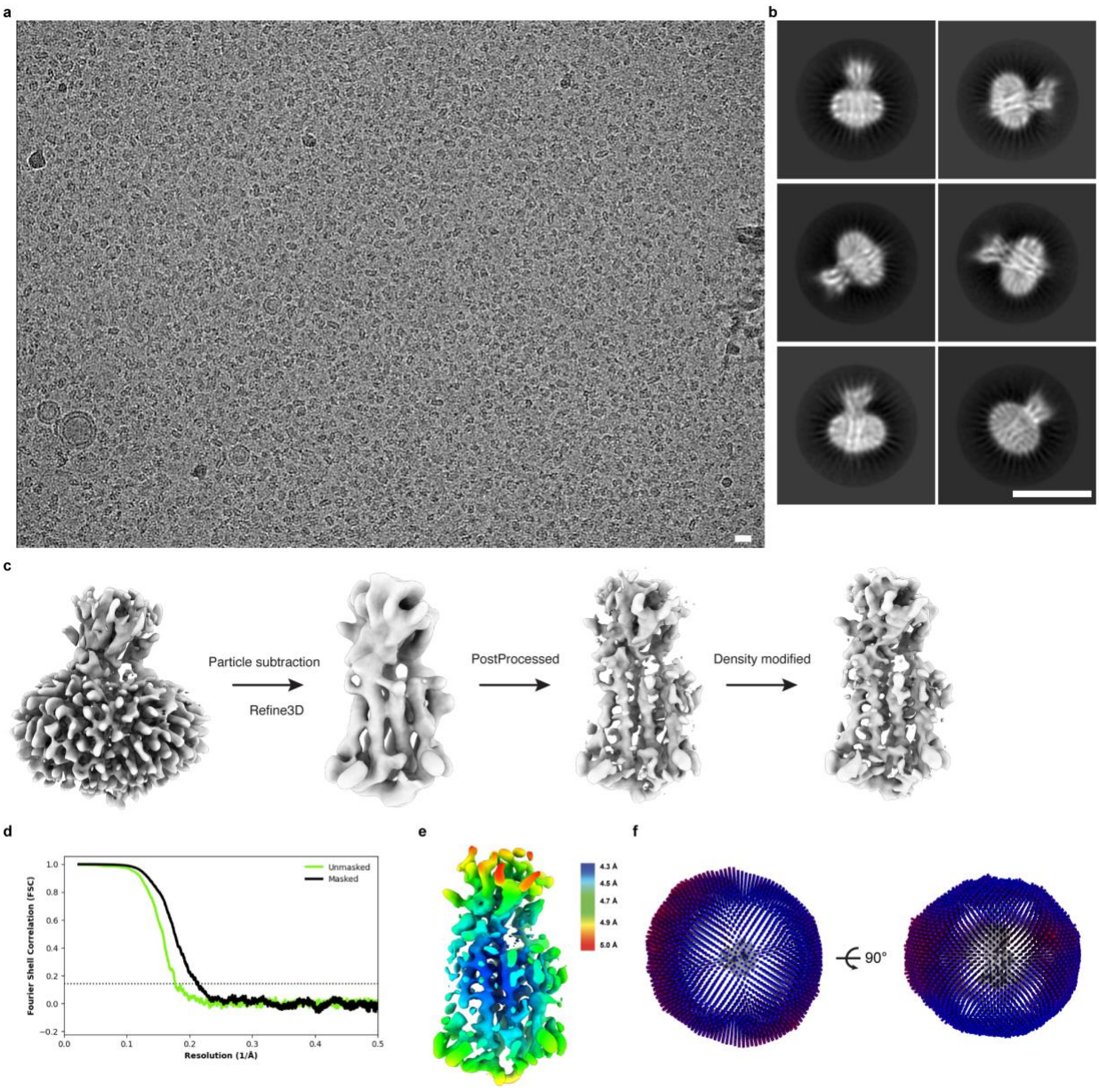
Refinement	
Model resolution (Å, FSC = 0.143 / FSC = 0.5)	3.9 / 5.9
Map-sharpening B factor (Å^2)	-421.5
Composition	
Number of atoms	3312
Number of protein residues	401
Number of ligands	1
R.m.s. deviations	
Bond lengths (Å)	0.004
Bond angles (Å)	0.855
Validation	
MolProbity score	1.64
Clashscore	5.98
Ramachandran plot	
Favored (%)	95.44
Allowed (%)	4.56
Disallowed (%)	0
Rotamer outliers (%)	0
Mean B factor (Å^2)	
Protein	153.9
Ligand	158.59

Table 4.1: Cryo-EM data collection, refinement, and validation statistics.

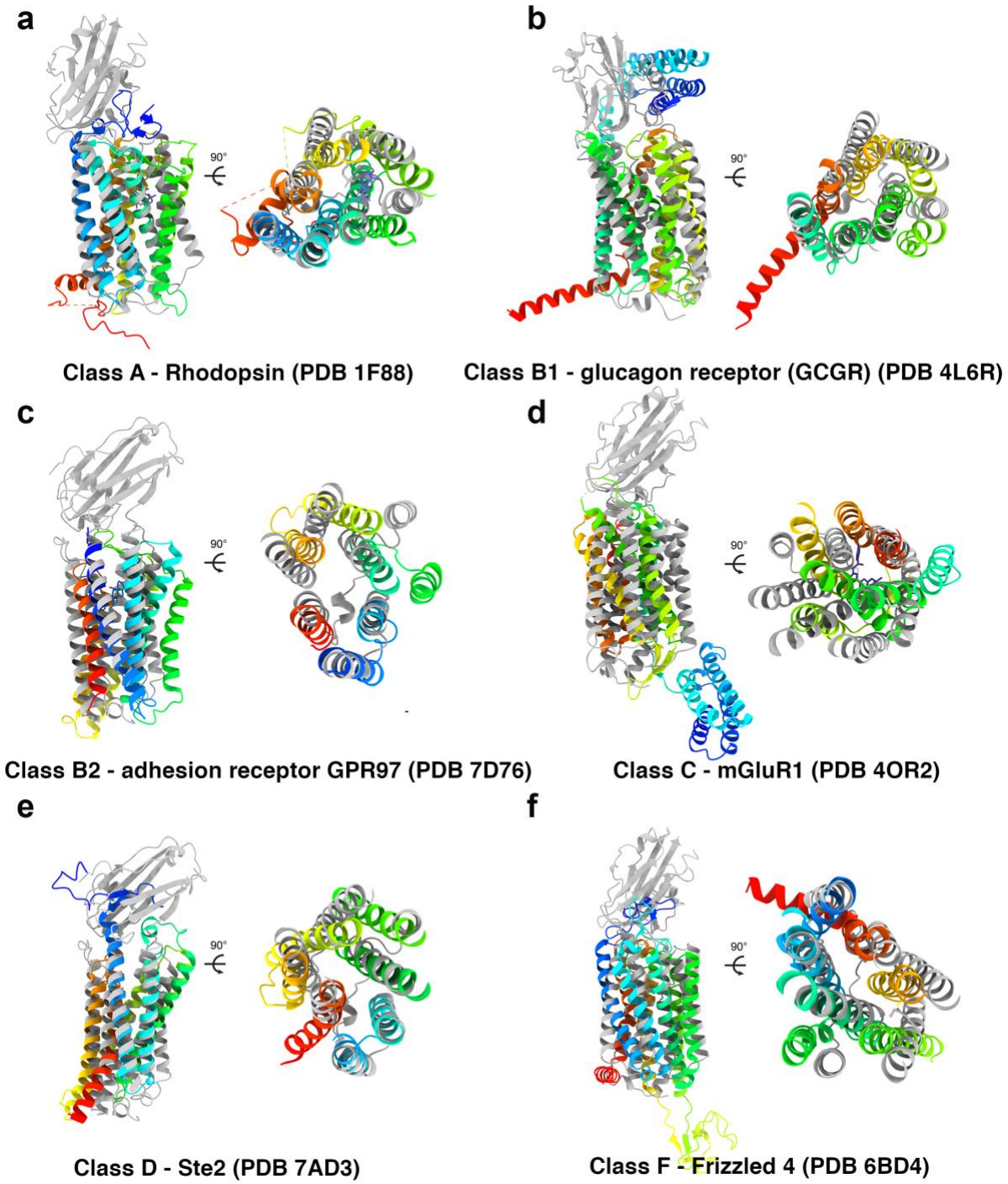
4.11 Supplemental figures



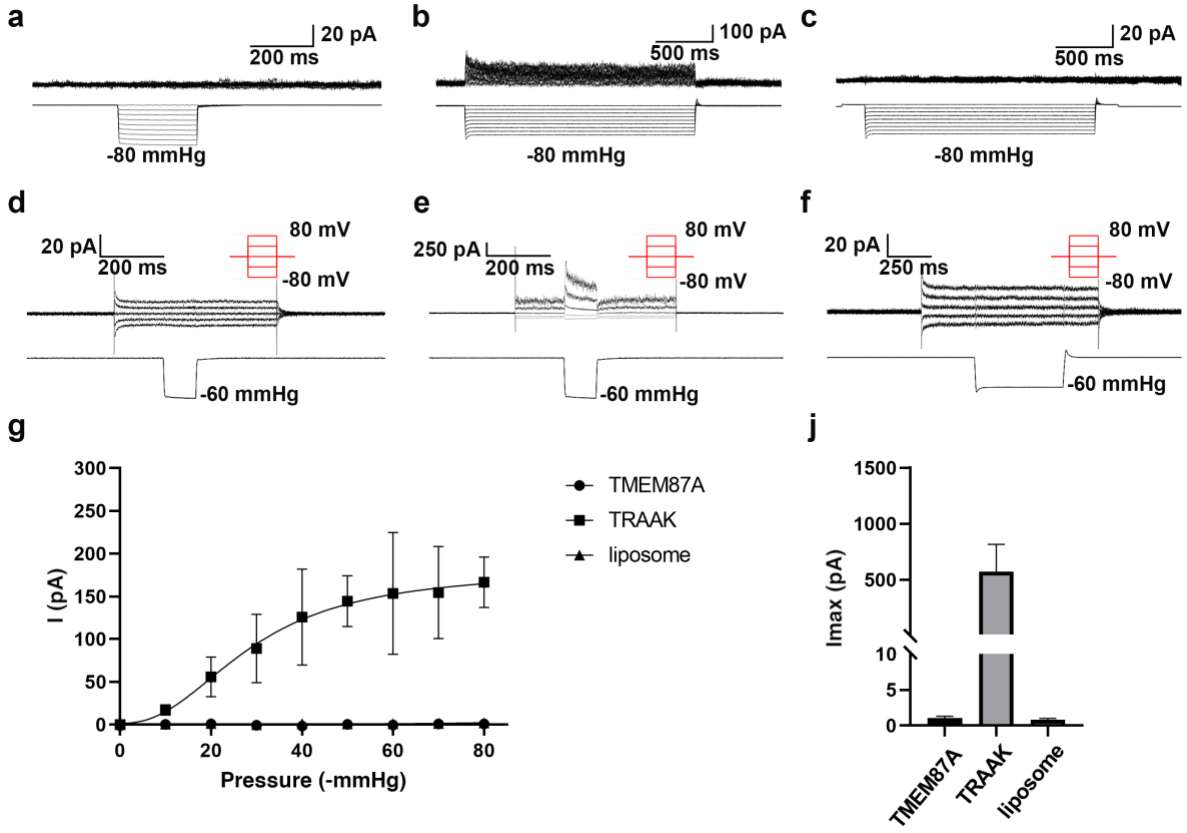
Supplemental Figure 4.1: TMEM87A biochemistry. **a**, Size-exclusion chromatogram and Coomassie stained SDS-PAGE gel showing the purification and reconstitution of TMEM87A in lipid nanodiscs. Shaded region corresponds to pooled fractions collected. Samples were run on a Superose 6 column.



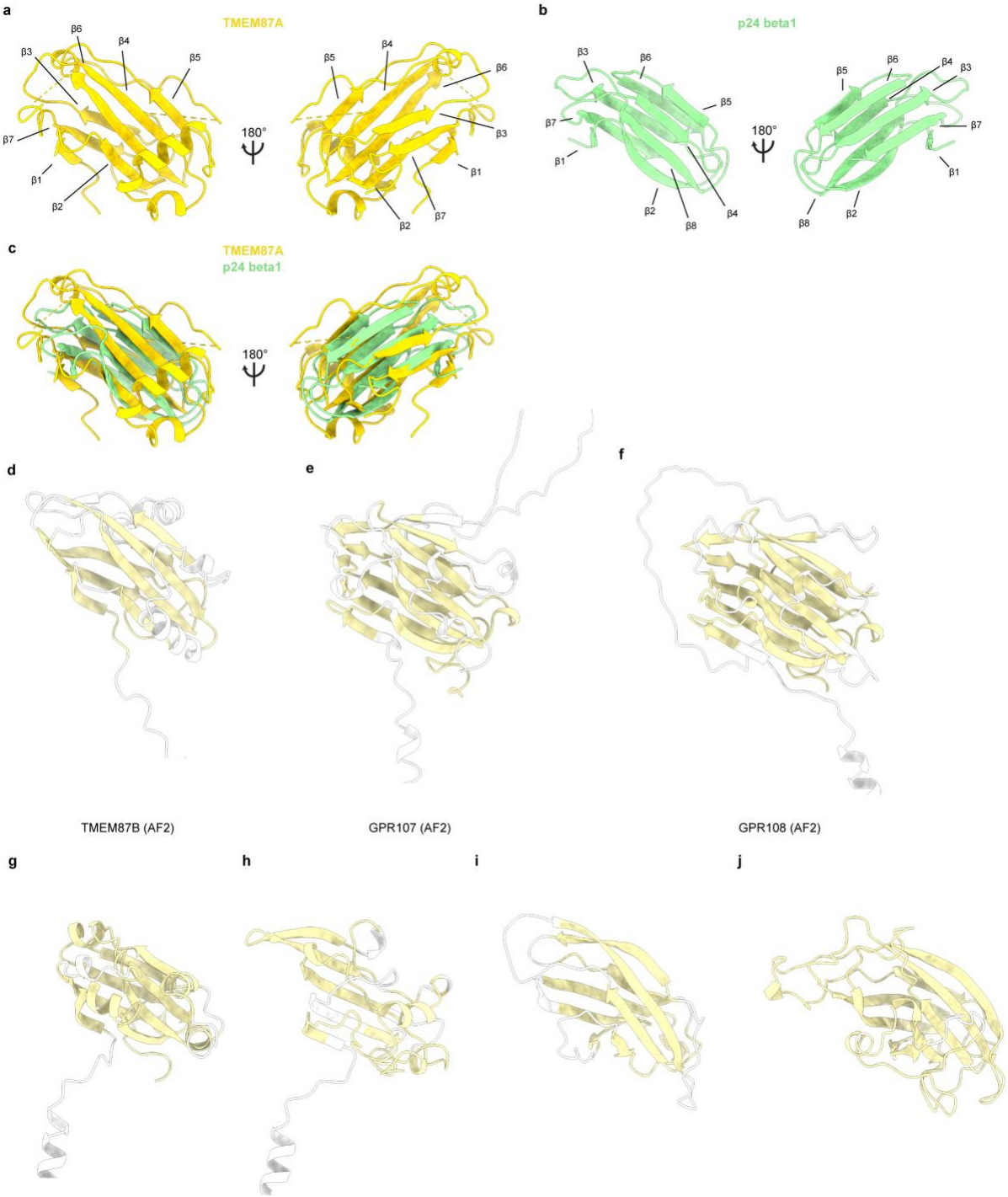
Supplemental Figure 4.2: Cryo-EM data, processing, and validation. a, Representative micrograph with 150 Å scale bar. **b,** Selected 2D class averages with 100 Å scale bar. **c,** Visual representation of final stages of the processing pipeline. **d,** Fourier shell correlations between the two corrected and masked (black) and unmasked (green) half maps. **e,** Relion-estimated local resolution colored on the cryo-EM map in rainbow. **f,** Angular distributions for particles used in the final refinement.



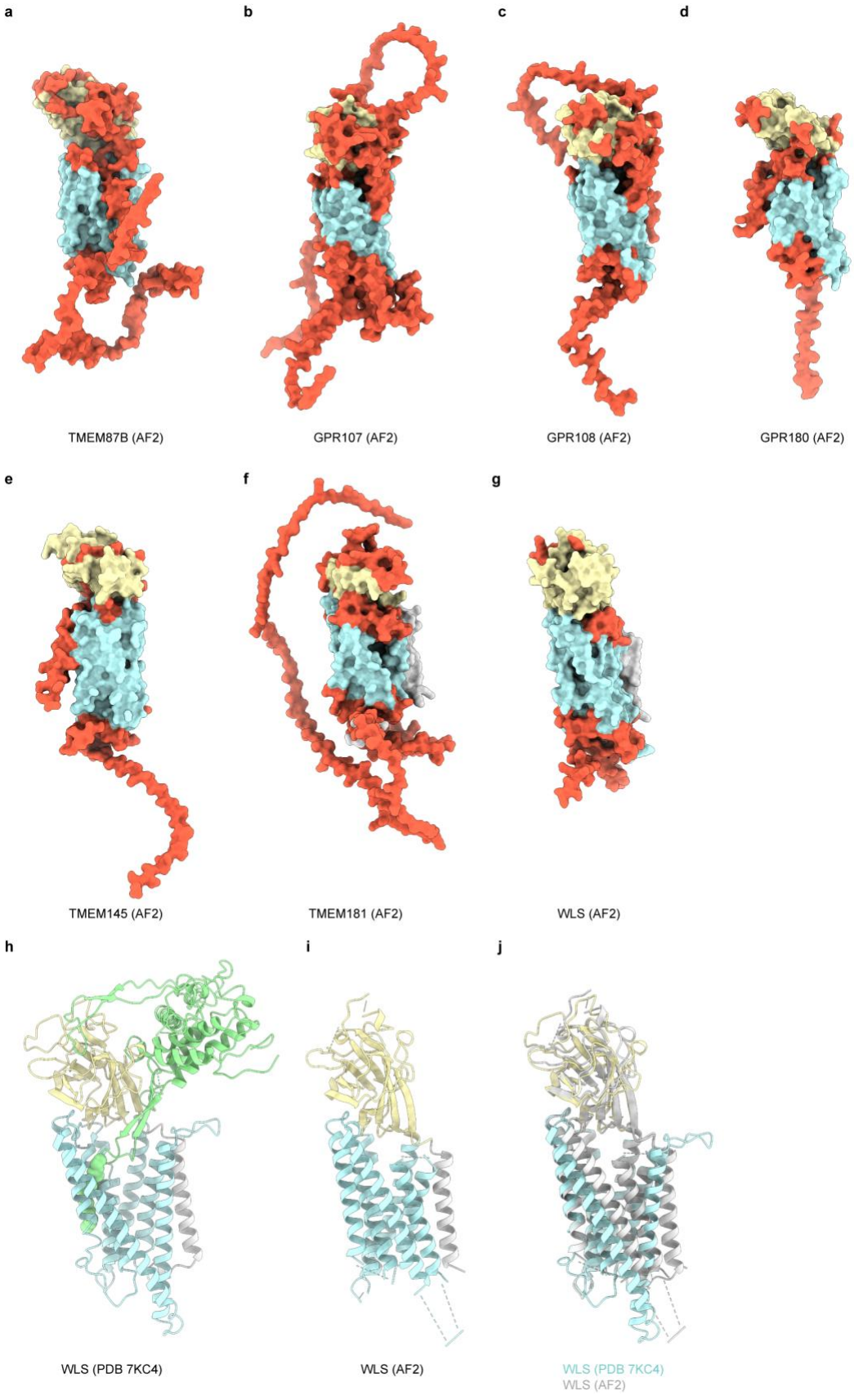
Supplemental Figure 4.3: Comparison of TMEM87A and GPCR structures. a-f Overlay of TMEM87A with multiple different GPCR structures, viewed from the membrane plane and 90° rotated view. **a,** Class A GPCR bovine Rhodopsin (PDB 1F88) **b**, Class B1 GPCR human glucagon receptor (GCGR) (PDB 4L6R) **c**, Class B2 GPCR human adhesion receptor GPR97 (PDB 7D76) **d,** Class C GPCR human metabotropic glutamate receptor 1 (mGluR1) (PDB 4OR2) **e,** Class D GPCR yeast Ste2 (PDB 7AD3) **f,** Class F GPCR human Frizzled 4 (PDB 6BD4). All GPCR structures colored rainbow from N-terminus (blue) to C-terminus (red), with TMEM87A colored in gray.



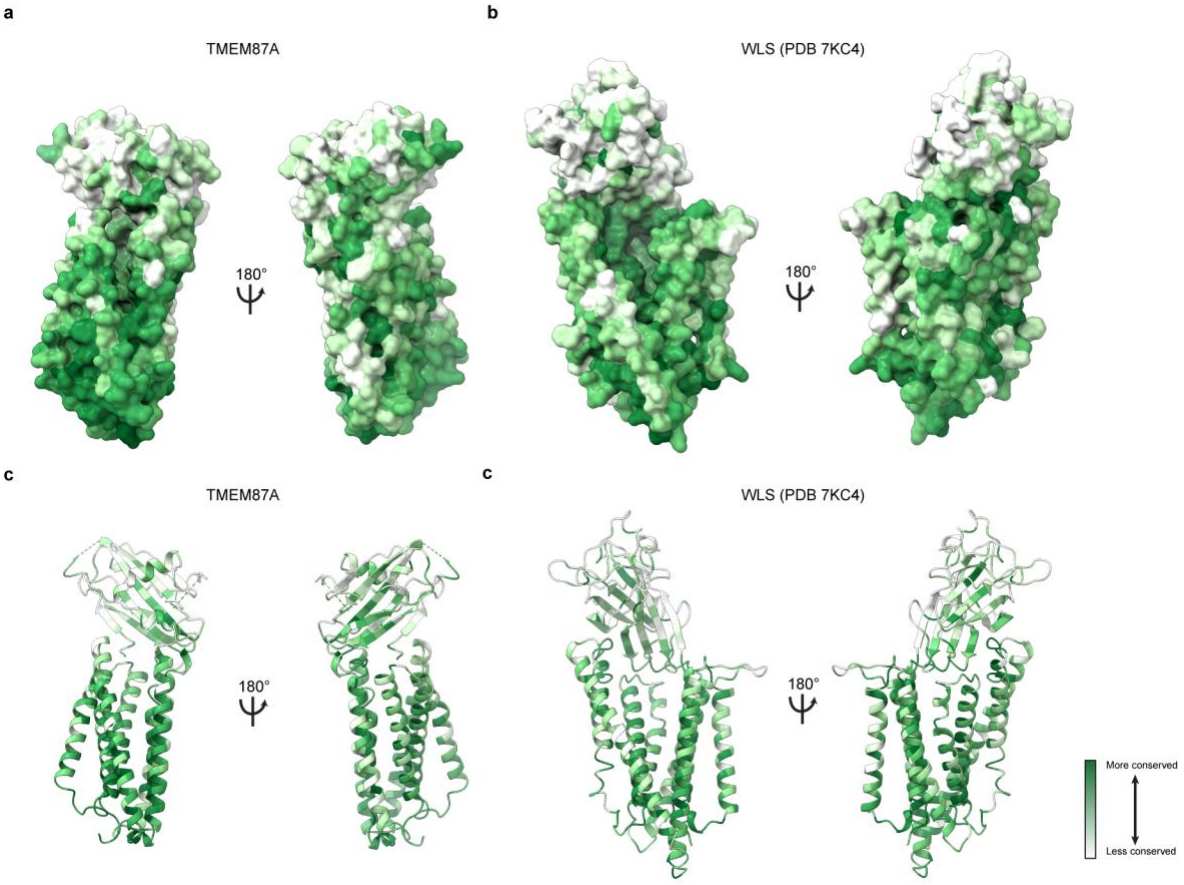
Supplemental Figure 4.4: TMEM87A proteoliposome recordings. a-c, Recordings of TMEM87A, TRAAK, and empty control proteoliposomes respectively with the application of increasing steps of pressure at constant membrane voltage of 60 mV. **d-f,** Recordings of TMEM87A, TRAAK, and empty control proteoliposomes respectively with a single application of constant pressure at -60 mmHg, and increasing steps of membrane voltage. **g,** Current / pressure plot for TMEM87A (circles), TRAAK (squares), and empty control proteoliposomes (triangles). **h,** Plot of maximum current observed for TMEM87A, TRAAK, and empty control proteoliposomes.



Supplemental Figure 4.5: GOLD domain comparisons. a, GOLD domain from present TMEM87A structure. **b,** p24 beta1 GOLD domain (PDB 5AZW) colored light green. **c** Overlay of TMEM87A and p24 GOLD domains. **d-j,** GOLD domains from AlphaFold2 predicted structures for TMEM87B, GPR107, GPR108, GPR180, TMEM145, GPR181, and WLS respectively, with GOLD domain colored light yellow and regions of low confidence (< 80 pLDDT) colored white.


a
b
c
d
TMEM87B (AF2)
GPR107 (AF2)
GPR108 (AF2)
GPR180 (AF2)
e
f
g
TMEM145 (AF2)
TMEM181 (AF2)
WLS (AF2)
h
i
j
WLS (PDB 7KC4)
WLS (AF2)
WLS (PDB 7KC4)
WLS (AF2)

Supplemental Figure 4.6: AlphaFold2 predictions of GOST family protein structures. **a-g,** Surface representations of the AlphaFold2 predicted structures for TMEM87B, GPR107, GPR108, GPR180, TMEM145, GPR181, and WLS respectively. GOLD domains are colored in light yellow, seven-transmembrane domains colored in light blue, N terminal helix colored in light gray (if applicable), and regions of low confidence (< 80 pLDDT) colored orange. **h,** Experimental WLS structure (7KC4) with GOLD domain colored light yellow, transmembrane domain colored light blue, N terminal helix colored light gray and Wnt8a colored light green. **i,** AlphaFold2 predicted WLS structure (only high confidence residues with pLDDT >=80 shown) with GOLD domain colored light yellow, transmembrane domain colored light blue, and N terminal helix colored light gray. **j,** Overlay of experimental and predicted WLS structures, with experimental structure (PDB 7KC4) colored and AlphaFold2 predicted structure in gray.



Supplemental Figure 4.7: Conservation of TMEM87A and WLS. a,c, TMEM87A structure colored according to conservation of each residue, from white (less conserved) to dark green (more conserved), on surface and ribbon representations respectively. **b, d,** WLS structure colored according to conservation of each residue, from white (less conserved) to dark green (more conserved), on surface and ribbon representations respectively.

Chapter 5 Concluding remarks

5.1 The study of small membrane protein structure and function by cryo-EM

Ion channels as a family of proteins contain tremendous functional diversity, and while substantial effort has led to a deep mechanistic understanding of some channels, many others remain poorly understood or entirely uncharacterized.[1] Additionally, new channels are discovered regularly and multiple currents have been characterized for which the molecular constituents are unknown.[2,3] The combination of structural methods and reductionist functional experiments is a valuable approach for understanding ion channel molecular mechanism, and is especially useful in cases where the pore-forming subunit has been unclear or debated. Here, I have shown the application of this approach, with particular focus on the use of cryo-EM to determine small membrane protein structures, to study three putative ion channels. Taken together, this work provides additional evidence for the value of this approach and has also provided a deeper understanding of each of these proteins function. Finally, it has established an approach for the determination of unmodified small membrane protein (<100 kDa) structures by cryo-EM.

Intriguingly, in only one of these three cases – for ORF3a – did we find compelling evidence for the protein functioning as an ion channel. In the case of TTYH family proteins, we argue that the family does not act as ion channels, and for TMEM87A, while not determined conclusively, it seems more likely that the protein does not function as an ion channel. For both TTYH proteins and for TMEM87A, these results are contrary to prior understanding in the literature, and demonstrate the importance of such a biophysical and reductionist approach to understand ion channel mechanism.

5.2 3a

For ORF3a, our work provided not only the first identification of SARS-CoV-2 3a ion channel activity, but also provided the first structure of any 3a protein, and the discovery of the tetrameric species. Our work was able to rule out emodin as a potential inhibitor, and characterized the inhibitory effect of ruthenium red. Given the constraints and importance of the ongoing COVID-19 pandemic, it is worth noting how the use of cryo-EM allowed for the rapid determination of these structures, with the project moving from conception to the release of an initial preprint in approximately 3 months.

Despite this increased biophysical understanding of 3a, many questions remain. Perhaps most significant is a lack of understanding of 3a's role in SARS-CoV-2 infection and the progression of COVID-19. The determination of an open-state structure of 3a will be critical for understanding 3a's mechanism, and will permit an understanding of the ion permeation pathway(s), key residues, and more. Future work will also uncover the role of the 3a tetramer and a better understanding of the evolutionarily related M protein, which is also predicted to have a similar fold as 3a.[4,5,6]

5.3 TTYH

In contrast, our study of TTYH family proteins provided strong evidence to suggest that TTYHs are not pore-forming subunits, and do not function as ion channels. The determination of TTYH

structures by cryo-EM was critical to this work for multiple reasons. First, it revealed the architecture of the TTYH family and the lack of any obvious permeation pathway. Second, it provides an informed context from which to examine previous mutations suggested to impact TTYH "channel" function, revealing that all of these mutations were unlikely to be directly related to TTYH functioning as an ion channel. Third, it allowed for the startling discovery of the TTYH *trans*-dimer. As a whole, this work allowed our refutation of TTYH's function as an ion channel and the proposal of a model in which TTYH acts through binding to and regulation of integrins.

In addition to testing the hypothesized function with integrins there are many other open questions for future work studying TTYH function. Perhaps most significant is an understanding of the physiological contexts of the *trans*-dimer, and its biological function. A major component will be a deeper understanding of the regulation of oligomeric state by calcium concentration. What precisely is the necessary concentration threshold of calcium to mediate this transition, as our structures represent snapshots at dramatically different concentrations? And accordingly, which physiological environments experience such a change in calcium concentration? Additionally, understanding how TTYHs regulate volume-regulated anion currents is a topic of interest. For example, could the regulation of these currents due to some direct interaction with TTYH or downstream of TTYH's regulation of integrin signaling? Finally, it will be important to leverage greater understanding of TTYH's function in the context of disease, such as glioblastoma.

5.4 TMEM87A

With TMEM87A, while the initial structure and proteoliposome recording data suggest that TMEM87A is unlikely to function as an ion channel, they do not conclusively eliminate the possibility. Intriguingly, this structure revealed the presence of TMEM87A's Golgi dynamics (GOLD) domain, facilitating the discovery of the larger family of GOLD domain seven-transmembrane proteins. This protein family appears to have a likely role in the regulation of protein transport to and from the Golgi, and perhaps act as "coincidence detectors" in a manner analogous to the function of WLS, which has evolved to simultaneously recognize aspects of Wnt proteins and their posttranslational modifications. Such a mechanism could suggest that TMEM87A and other family members play a key role in the transport of lipidated secreted molecules such as Wnts, ghrelin, cytokines, and more.

Future work with TMEM87A will investigate potential cargo binding through both functional and structural methods. A better understanding of TMEM87A regulation of protein transport could also inform its potential role in the recently described mechanotransduction pathway. If TMEM87A is not the pore-forming subunit of this complex, then how is it impacting this current? Could it be through regulation of the transport of the channel itself? Or perhaps TMEM87A transports a signaling molecule which impacts the expression of this channel. Future work will provide a better understanding, not only of TMEM87A but of the GOLD domain seven-transmembrane protein family as a whole. Of particular interest are experiments testing the hypothesis that GOLD domain seven-transmembrane proteins other than WLS might act to transport lipidated secreted proteins such as Wnt. One preliminary experiment would be to investigate the potential binding of Wnt family members against the family of GOLD domain

seven-transmembrane proteins to see if family members other than WLS can transport Wnts. Additionally, it will be of significant interest to investigate what other molecules could comprise potential cargo, perhaps through pull-down experiments with each family member, ideally from its endogenous context.

5.5 Future developments in membrane protein structural biology

Current developments in cryo-EM will allow even greater insights in ion channel structure and function in the future. In particular, three areas of research are worth noting: the imaging of reconstituted proteoliposomes by cryo-EM, the imaging of ion channels in their native environment by cryo-electron tomography (cryo-ET), and improvements in core technology such as the development of laser phase plates, improved computational methods, improved sample preparation substrates, and more.

The use of reconstituted proteoliposomes provides an experimental system with a reconstituted bilayer of desired composition akin to lipid nanodiscs, but with the ability to independently manipulate the internal and external solutions unlike a lipid nanodisc.[7,8] This mechanic has been suggested to allow for the visualization of states previously inaccessible, such as an ion channel experiencing membrane tension due to cell swelling, or manipulation of the membrane potential of the proteoliposome.[7,8] While the method has been demonstrated successfully with large symmetric channels, an improvement in either throughput or signal-to-noise will be necessary for the approach to be more widely useful.[9,10,11,12]

Cryo-ET provides a method for the visualization of macromolecules within a cell in their native environment.[13] Currently very technically challenging, yet in some cases has yielded stunning results, including the visualization of an antibiotic bound ribosome to 3.5 Å within a cell.[13,14] However, for membrane proteins, results have thus far been much more limited.[15,16,17] As membrane proteins are smaller, often expressed at lower levels, and embedded in the membrane, they present a greater technical challenge than many of the macromolecules which have been imaged successfully to date. Given the rapid pace of technological improvement in the field, especially with the potential for transformative improvements such as a laser phase plate, the study of membrane proteins by cryo-ET will likely become more tractable.

One of the most significant technical barriers in cryo-EM is the very low signal-to-noise, and many technical developments are focused on improving signal while reducing noise. Improvements in substrate stability, new technologies like laser phase plates, and improved computation all seek to improve signal-to-noise.[18,19,20,21,22] These benefits will be useful not only for single particle cryo-EM of purified proteins but also to make more challenging methods like imaging of proteoliposomes and cryo-ET, more tractable. These improvements will directly translate to a better understanding protein structure and function, especially for cases which are currently challenging, such as small membrane proteins.

Beyond advances specific to cryo-EM, one of the most transformative developments in structural biology has been the improvement in protein structure prediction demonstrated by DeepMind's AlphaFold2 and related RoseTTAFold.[5,23] The impact of AlphaFold2 has been significant and should not be understated – not only a paradigm shift for the field of protein structure prediction,

but with impacts in structural biology (such as using predictions for the phasing of crystals, docking into low resolution cryo-EM maps, construct design, and more) and beyond.[24] The impact of AlphaFold2 is clearly demonstrated in the above study of TMEM87A. In this case not only did AlphaFold2 provide a high quality model which could be used as a foundation for building a final protein model in the cryo-EM map, but in conjunction with bioinformatics tools such as Dali allowed critical novel insights leading to the discovery of the larger GOLD domain seven-transmembrane family of proteins, and the proposed hypothesis for their function.[25] Intriguingly, when a similar approach is applied for TTYH, it uncovers the protein Prominin-1 as having a similar fold, which had recently been hypothesized to be structurally similar.[26] The use of AlphaFold2 to identify proteins, without experimentally determined structures, which may share a fold with a protein of interest, is just one of many ways that AlphaFold2 has impacted modern structural biology. As both structural biology and artificial intelligence continue to develop, the two fields will likely to continue to positively impact each other.[27]

5.6 Conclusion

Taken together, these results have demonstrated the value of using structural and functional "bottom-up" methods to study poorly characterized putative ion channels. In the case of each protein studied, these results provide a foundation for future work pursing deeper mechanistic understanding of each protein. From a technical perspective, this work has pushed the boundary of what was considered possible in single particle cryo-EM, demonstrating the capacity to determine the structure of full-length, unmodified small (< 100 kDa) membrane proteins without the use of Fab or other added mass. In the case of 3a, the apparent benefit from the use of a cutting-edge energy filter provided, to the best of our knowledge, the current highest resolution cryo-EM structure of a small membrane protein (< 100 kDa) thus far at ~2.1 Å. As technology continues to develop, the benefit of cryo-EM for the study of ion channel structure and function will likely continue to improve.

5.7 References

1. Hille B. Ionic channels in excitable membranes. Current problems and biophysical approaches. *Biophysical Journal.* **22**(2), 283-94 (1978).
2. Yang, J. *et al.* PAC, an evolutionarily conserved membrane protein, is a proton-activated chloride channel. *Science* **364**, 395–399 (2019).
3. Hill, R. Z. & Bautista, D. M. Getting in Touch with Mechanical Pain Mechanisms. *Trends in Neurosciences* **43**, 311–325 (2020).
4. Tan, Y. *et al.* Unification and extensive diversification of M/Orf3-related ion channel proteins in coronaviruses and other nidoviruses. *Virus Evolution* **7**, (2021).
5. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
6. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* **50**, D439–D444 (2022).
7. Wang, L. & Sigworth, F. J. Structure of the BK potassium channel in a lipid membrane from electron cryomicroscopy. *Nature* **461**, 292–295 (2009).

8. Sejwal, K. *et al.* Proteoliposomes – a system to study membrane proteins under buffer gradients by cryo-EM. *Nanotechnology Reviews* **6**, 57–74 (2017).
9. Tonggu, L. & Wang, L. Broken symmetry in the human BK channel. *bioRxiv* (2018).
10. Yao, X., Fan, X. & Yan, N. Cryo-EM analysis of a membrane protein embedded in the liposome. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 18497–18503 (2020).
11. Melville, Z., Kim, K., Clarke, O. B. & Marks, A. R. High-resolution structure of the membrane-embedded skeletal muscle ryanodine receptor. *Structure* **30**, 172-180.e3 (2022).
12. Yang, X. *et al.* Structure deformation and curvature sensing of PIEZO1 in lipid membranes. *Nature* **604**, 377–383 (2022).
13. Turk, M. & Baumeister, W. The promise and the challenges of cryo-electron tomography. *FEBS Lett* **594**, 3243–3261 (2020).
14. Tegunov, D., Xue, L., Dienemann, C., Cramer, P. & Mahamid, J. Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells. *Nat Methods* **18**, 186–193 (2021).
15. Gold, V. A. M. *et al.* Visualizing active membrane protein complexes by electron cryotomography. *Nat Commun* **5**, 4129 (2014).
16. Schaffer, M. *et al.* Optimized cryo-focused ion beam sample preparation aimed at in situ structural studies of membrane proteins. *Journal of Structural Biology* **197**, 73–82 (2017).
17. Lamm, L. *et al.* MemBrain: A Deep Learning-aided Pipeline for Automated Detection of Membrane Proteins in Cryo-electron Tomograms. *bioRxiv*, (2022). doi:10.1101/2022.03.01.480844.
18. Schwartz, O. *et al.* Laser phase plate for transmission electron microscopy. *Nat Methods* **16**, 1016–1020 (2019).
19. Zhong, E. D., Lerer, A., Davis, J. H. & Berger, B. Exploring generative atomic models in cryo-EM reconstruction. *arXiv* (2021).
20. Rosenbaum, D. *et al.* Inferring a Continuous Distribution of Atom Coordinates from Cryo-EM Images using VAEs. *arXiv* (2021).
21. Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat Methods* **18**, 176–185 (2021).
22. Punjani, A. & Fleet, D. J. 3D Flexible Refinement: Structure and Motion of Flexible Proteins from Cryo-EM. *bioRxiv,* (2021). doi:10.1101/2021.04.22.440893.
23. Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
24. Cramer, P. AlphaFold2 and the future of structural biology. *Nat Struct Mol Biol* **28**, 704–705 (2021).
25. Holm, L., and Sander, C. Dali: a network tool for protein structure comparison. *Trends in Biochemical Sciences* **20** (11), 478-480 (1995).
26. Hori, A. *et al.* Prominin-1 Modulates Rho/ROCK-Mediated Membrane Morphology and Calcium-Dependent Intracellular Chloride Flux. *Sci Rep* **9**, 15911 (2019).
27. AlQuraishi, M. Machine learning in protein structure prediction. *Current Opinion in Chemical Biology* **65**, 1–8 (2021).